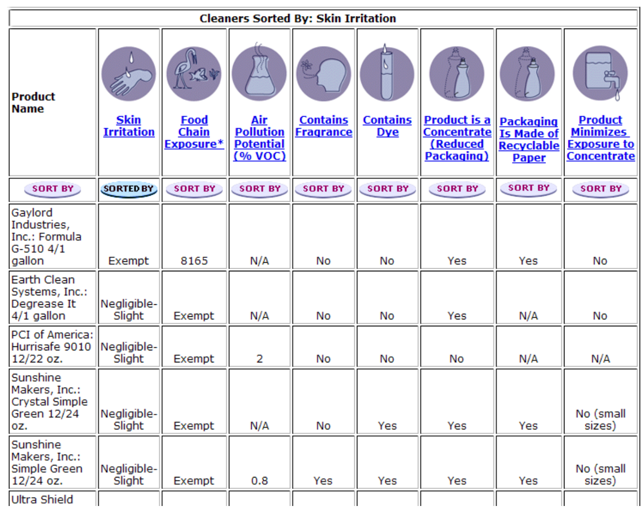
Personalizing User Interfaces for Environmental Decision Support Systems

Timothy Maciag, Daryl H. Hepting, Dominik Slezak
Department of Computer Science, University of Regina
3737 Wascana Parkway, Regina, Saskatchewan, S4S 0A2, Canada
Email: {maciagt, hepting, slezak}@cs.uregina.ca

Abstract

The quality of the natural environment has become one of the primary concerns in present society. In Canada, we have been asked to take on the "One Tonne Challenge" to reduce personal household emissions by 1 tonne. However, very little has been done to illuminate the various connections between our household purchases and the effect they can have on the quality of our health and environment. Several decision support systems are available to assist consumers compare alternatives. However, these systems do little to enhance the consumer's experience. Correct clustering of consumers in terms of their product attribute preferences would enable the construction of personalized user interfaces thus increase consumer satisfaction when interacting with the system and increase the chance of inspiring greener purchasing habits. This paper analyzes a clustering technique that uses methods from multivariate statistics, rough set theory, and machine learning to cluster users in a web-based environmental decision support system and test the success of the clustering. Results from our analysis are discussed.



Cleaners Sorted By: Skin Irritation									
Product Name	Skin Irritation	Food Chain Exposure*	Air Pollution Potential (% VOC)	Contains Fragrance	Contains Dye	Product is a Concentrate (Reduced Packaging)	Packaging Is Made of Recyclable Paper	Product Minimizes Exposure to Concentrate
SORT BY	SORTED BY	SORT BY	SORT BY	SORT BY	SORT BY	SORT BY	SORT BY	SORT BY
Gaylord Industries, Inc.: Formula G-510 4/1 gallon	Exempt	8165	N/A	No	No	Yes	Yes	No
Earth Clean Systems, Inc.: Degrease It 4/1 gallon	Negligible-Slight	Exempt	N/A	No	No	Yes	N/A	No
PCI of America: Hurrisafe 9010 12/22 oz.	Negligible-Slight	Exempt	2	No	No	No	N/A	N/A
Sunshine Makers, Inc.: Crystal Simple Green 12/24 oz.	Negligible-Slight	Exempt	N/A	No	Yes	Yes	Yes	No (small sizes)
Sunshine Makers, Inc.: Simple Green 12/24 oz.	Negligible-Slight	Exempt	0.8	Yes	Yes	Yes	Yes	No (small sizes)
Ultra Shield								

Figure 1. A Screen Capture of the US-EPA EPP wizard that compares cleaning products. The wizard tool enables users to sort cleaning products using 3 different tools. This screen capture is 1 of those tools.

1 Introduction

The Kyoto accord for reduction of greenhouse gas (GHG) emissions represents a step toward a cleaner environment. Although Canadians have been asked to take on the "One Tonne Challenge" to reduce personal household emissions by 1 tonne, very little has been done to illuminate the various connections between our household purchases and the effect they can have on the quality of our health and environment. The environmental soundness of many of these purchases goes unquestioned. If we consider a life cycle assessment approach, it becomes clear that many of these purchases have potential to create a negative impact, however indirectly.

Several environmental decision support systems (EDSS) are available to assist consumers compare alternatives. However, these systems seldom attempt to enhance the consumer's experiences when interacting with the system. Correct clustering of consumers in terms of their product attribute preferences would allow the construction of personalized user interfaces and increase consumer satisfaction by emphasizing those items with attributes that meet their specific values. By allowing consumers to also compare their current choices with possible alternatives, some of which may be more eco-effective[1], the chance of inspiring greener

[1]The term eco-effective refers to the concept of producing and consuming items that have a positive impact on both our health

purchasing habits is great. A preliminary usability study of a web-based EDSS, based on the United States Environmental Protection Agency's (US-EPA) environmentally preferable purchasing (EPP) wizard that compares cleaning products[2], was conducted. The system interface is illustrated in Figure 1.

Eight environmental and health related system features were used to distinguish between cleaning products. These included:

1. **Skin Irritation (skin):** Refers to the presence of chemicals in the cleaning product that cause redness or swelling of skin. Attribute values range from the most preferable to least preferable value, i.e. negligible, slight, moderate, or strong. A special skin irritation value of "exempt" signifies that there is less than 5% (by weight) chemical component in the product.

2. **Food Chain Exposure (fce):** Refers to ingredients in cleaning products that have the potential to be introduced into the food chain by being consumed by smaller aquatic plants and animals which are than consumed by larger animals. Food chain exposure is measured by calculating a products bioconcentration factor (BCF). Products with a BCF less than 1000 or a BCF of "exempt" are more preferable.

3. **Air Pollution Potential (air):** Refers to products that may contain volatile organic compounds (VOC), i.e. compounds that have the potential to form atmospheric pollutants, e.g. smog. These pollutants can cause eye, nose, throat, and lung irritation, as well as trigger asthma attacks. The lower the VOC, the more preferable the product with a special value of "N/A" (not applicable), which indicates that there are no VOCs of concern present, being the most preferable.

4. **Product Contains Fragrances (frag):** Refers to fragrances that are added to the cleaning product to improve, or mask, its "natural" odor.

5. **Product Contains Dye (dye):** Refers to dyes that have been added to the cleaning product to change the color of the product.

6. **Product uses Recyclable Packaging (rec):** Refers to cleaning products that are packaged using recyclable packaging.

7. **Product is a Concentrate (con):** Refers to cleaning products that are packaged using reduced packaging, e.g. packaging the cleaning product in a recyclable plastic bag which acts as a refiller for use in its original packaging.

8. **Product Reduces Exposure to Concentrate (exp):** Refers to cleaning products that reduce exposure to concentrated packaging. Concentrates have potential to increased healths as it may place the consumer at greater exposure to potentially hazardous ingredients of the product.

48 University of Regina undergraduate students were recruited through the University of Regina participant pool[3]. Participants were asked to perform a variety of tasks to test the usability of the system. As well, participants were asked to complete pre-task and post-task questionnaires relating their technical experience as well as their experiences while interacting with the system. Of the questions asked in the post-task questionnaire, participants were asked to rank the system features using a four point scaling, i.e. unimportant, somewhat important, important, and very important. One of the goals of our analysis was to cluster users based on their system feature preferences. Another goal of our analysis was to simplify the clustering process by observing the possibility of reducing the dimensions of the system features.

Our algorithm uses elements from multivariate statistics, rough set theory, and machine learning. Specifically, the k-means clustering algorithm was used to formulate the initial and future user clusters. Attribute reduction, as used in rough set theory, was used to reduce the dimensionality of the user ranked data. The success of the initial and future clustering was measured using a train and test procedure common in machine learning, specifically in unsupervised learning, wherein the the total population of participants and their associated ranked attributes were split into train and test sets thus enabling our algorithm to be adequately tested. Our algorithm and the results of our analysis are discussed in detail in this paper.

and the natural environment, e.g. a 100% biodegradable fabric, as opposed to an item that simply limits the impact on our health and the natural environment, e.g. a plastic recyclable bottle [9].

[2]US-EPA EPP wizard available online: `http://www.epa.gov/opptintr/epp/cleaners/select/matrix.htm`. The system is only partially functional (Accessed June 2005).

[3]The University of Regina participant pool is a program available to researchers at the University of Regina which enables them to recruit participants for research and study purposes. Participants are usually University of Regina undergraduate students. As a result of their participation, students are rewarded a bonus grade in a participating computer science or psychology course of their choosing.

The structure of this paper is as follows. Section 2 will introduce the key concepts and techniques used to formulate and test the derived clustering methodology. Section 3 will discuss the motivation and procedures behind our analysis. Section 4 will discuss results discovered from a case study that tested the derived methodology and section 5 will state our conclusions and discuss related future work.

2 Key Concepts

This section describes the key concepts and techniques of our approach in terms of the origins in multivariate statistics, rough set theory, and machine learning.

2.1 Clustering

Clustering is a common technique of multivariate statistics [14]. The fundamental goal of clustering is to formulate "natural" groupings of similar data within a data set without any prior knowledge of the specific class designations of the data [8]. One of the goals of our analysis was to cluster EDSS users according to their perceived preferences of various health and environmental related attributes. The purpose of the clustering was to formulate groupings of users with similar attribute preferences so as to observe the potential to efficiently construct personalized EDSS user interfaces. These user interfaces would emphasize items containing attributes that meet the users' specific values according to their assigned cluster. There exist many clustering algorithms that could be used for such an analysis. The k-means clustering algorithm was chosen in our analysis due to its popularity and ease of use [12].

Clustering algorithms can be roughly divided into 2 main categories, *hierarchical clustering* and *partitional clustering* [8]. The k-means clustering algorithm functions as a partitional clustering algorithm. A variation of the k-means clustering algorithm as used in our analysis is described below: [12]

1. Select a set of features (attributes) to be clustered.
2. Select an appropriate value of k, i.e. the number of classes or partitions to formulate.
3. Randomly select k initial cluster centres using the Euclidean distance function.
4. Assign features a class designation according to the closeness to the nearest cluster centre.
5. Place the cluster centres in the centroid, or centre of mass, within each class partition
6. Repeat from *4)* until all features (attributes) are assigned to appropriate classes or partitions.

2.2 Rough Set Theory

Rough set theory, developed by Z. Pawlak in the early 1980's, provides techniques for representing uncertainty in knowledge systems [13]. The techniques of rough set theory enables the conceptualization of approximations based on feature classes [1]. One of the fundamental concepts of rough set theory is the simplification of knowledge representation systems, i.e. knowledge reduction [10].

Concepts of knowledge reduction include *reduct* and *core* formulation. A "reduct of knowledge" refers to the features necessary in the data to discern the objects in the classes, in our case user clusters, whereas a core refers to those features within reducts that are common in every reduct [10]. Data is visualized in the form of an *information table* or *decision table* [11]. Table rows represent system objects and decision attributes whereas the table columns represent the system features and decision attribute definition(s). A *lower approximation* is sought to distinguish those objects belonging to specific decision classes without any indiscernability. The *union* of the lower approximations of the decision class(es) is referred to as the *positive region* of the decision table. The use of rough set techniques as described above were used in our analysis to reduce the knowledge base by filtering only those features needed to discern objects, in our case, the users in the varying clusters.

Feature selection is a domain that has acquired a tremendous amount of interest lately [7]. Many papers that discuss the concept of feature selection do so concerning data sets with a large number of dimensions [6]. However, feature selection can also apply to data sets containing few features, as the data set used in our analysis illustrates. Rough set theory provided the means to reduce the dimensions of the particular data set we used. However, further dimension filtering needed to be performed in order to choose meaningful features. This was accomplished by considering the user rankings of the system features. Our technique and algorithm is discussed later in this paper.

2.3 Unsupervised Learning

Clustering is an excellent example of an unsupervised learning technique [4]. When clustering in an unsupervised learning environment, feature classes are not known [5], as was the case in our analysis. The k-means clustering algorithm provided the means to acquire a decision variable. The cluster value assigned to particular users essentially became the decision variables.

Unsupervised learning is important in machine learning research due to its humanistic nature, i.e. having similarities to actual human learning processes and associated brain activity patterns [4]. We employed unsupervised learning techniques in our analysis to test the result of our algorithm. For our analysis we formulated an algorithm that utilized methods and techniques from multivariate statistics, rough set theory, and machine learning. The algorithm we formulated is described as follows:

1. Split the population into 2 samples. One sample represents the training set whereas the other sample, a test set.
2. In the training set, formulate the clusters using the k-means clustering algorithm, with $k = 2$.
3. Test for decision reducts.
   (a) Using the results from the clustering in the training set, formulate the reducts (if any) with the newly acquired cluster value as the decision variable.
4. Filter the resulting reducts by analyzing the cluster centres of attributes represented in each of the user clusters choosing those reducts with attributes users in either cluster ranked as *important* and *very important*.
5. Formulate clusters in the test set using the k-means clustering algorithm and only those attributes in the reducts that made it through the filter process as described above.
   (a) Test each reduct separately using the described train and test procedure.

When referring to our algorithm as described above we first randomly distribute the population of participants along with their accompanying system feature rankings into 2 samples with 24 users in each sample. One sample was randomly selected as a training set. Using the k-means clustering algorithm, as provided by the powerful statistical software application SPSS[4], user clusters in the training set were formulated. The k-means clustering algorithm in SPSS utilizes the Euclidean distance function.

We tested k-values of 2, 3, 4, and 5. In order to choose the *best* value of k, we analyzed those attributes users ranked as *important* and *very important* in each user cluster. This was accomplished by observing and recording the attributes in each user cluster that had a cluster centre of 3 or greater for each value of k. The results were as follows:

- k = 2
  - Cluster 1 (13 users): skin, air
  - Cluster 2 (11 users): air, rec, con, skin, fce, exp
- k = 3
  - Cluster 1 (8 users): skin, air
  - Cluster 2 (9 users): air, con, skin, rec, frag, fce, exp
  - Cluster 3 (7 users): skin, air, exp
- k = 4
  - Cluster 1 (5 users): skin, fce, air
  - Cluster 2 (4 users): skin, rec, air
  - Cluster 3 (6 users): skin, air
  - Cluster 4 (9 users): air, con, skin, rec, frag, fce
- k = 5
  - Cluster 1 (6 users): skin, air, fce
  - Cluster 2 (4 users): skin, rec, air
  - Cluster 3 (2 users): skin, exp, air, fce
  - Cluster 4 (10 users): air, skin, con, rec, frag, fce, exp
  - Cluster 5 (2 users): skin, frag, air

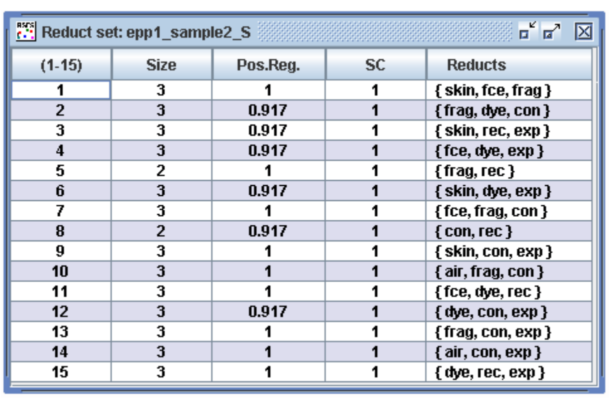
Based on the above results, we filtered the k-values as follows:

1. Observe whether the k-value gives a quality distribution of users in each cluster. This was accomplished by observing the distribution of users in each cluster for each value of k.

[4]Information on SPSS available online: `http://www.spss.com/` (Accessed July 2005)

2. Observe whether the k-value formulates distinct user clusters. This was accomplished by analyzing and comparing the attributes in each user cluster for each k-value and noting the similarities in attribute preferences (if any) among the user clusters.

When k = 2 we observe a quality distribution of users in each cluster. We also observe 2 distinct user clusters with varying user attribute preferences. When we analyze k = 3, we also observe a quality distribution of users in each cluster. However, we observe that users in clusters 1 and 3 have almost identical attribute preferences. When analyzing k = 4, we observe a satisfactory distribution of users in each user cluster. However, we again observe that users in different clusters have similar attribute preferences, as observed in clusters 1, 2, and 3. Finally, when analyzing k = 5, we observe a non-satisfactory distribution of users in each cluster. Thus, it was noted that testing for k-values above 5 was not necessary, as higher k-values would lead to non-satisfactory distribution of users in each cluster. Furthermore, when analyzing k = 5 we observe that users in clusters 1, 2, and 5 have similar attribute preferences. Therefore, we selected k = 2 as the most appropriate k-value for our analysis and thus clustered the users in the training set into 2 partitions.


Reduct set: epp1_sample2_S

(1-15)	Size	Pos.Reg.	SC	Reducts
1	3	1	1	{ skin, fce, frag }
2	3	0.917	1	{ frag, dye, con }
3	3	0.917	1	{ skin, rec, exp }
4	3	0.917	1	{ fce, dye, exp }
5	2	1	1	{ frag, rec }
6	3	0.917	1	{ skin, dye, exp }
7	3	1	1	{ fce, frag, con }
8	2	0.917	1	{ con, rec }
9	3	1	1	{ skin, con, exp }
10	3	1	1	{ air, frag, con }
11	3	1	1	{ fce, dye, rec }
12	3	0.917	1	{ dye, con, exp }
13	3	1	1	{ frag, con, exp }
14	3	1	1	{ air, con, exp }
15	3	1	1	{ dye, rec, exp }

Figure 2. Total set of reducts formulated from the training set using RSES v.2.2.

After performing the clustering on the training set we used the newly acquired user cluster values as the decision variable and formulated the decision reducts [2]. Using the Rough Set Exploration System (RSES)[5] we formulated the reducts. We decided to test our algorithm using approximate reducts therefore we shortened the reduct set using a 90% shortening ratio [3] thus achieving a 90% positive region in the training set decision table. 15 reducts, as illustrated in Figure 2, were formulated.

Next, we applied our filtering algorithm to the 15 reducts to include only those reducts containing attributes that the users ranked as *important* and *very important* by again observing the cluster centres of the attributes in each the user cluster, as depicted in Table 1. Recall, cluster centre values of 3 or greater include those attributes within the range desired. Reducts containing attributes with cluster centres of less than 3 in **both** clusters were omitted from further observation. By observing the results in Table 1 we see that the attributes *product contains fragrance (frag)* and *product contains dye (dye)* are both attributes below the desired observation range in **both** user clusters. Therefore, reducts containing those 2 attributes were omitted from further analysis.

Cluster 1	Cluster 2
skin (3.85)	air (3.91)
air (3.38)	rec (3.64)
fce (2.77)	con (3.64)
exp (2.15)	skin (3.55)
rec (2.00)	fce (3.09)
frag (1.92)	exp (3.09)
dye (1.85)	**frag (2.82)**
con (1.77)	**dye (2.18)**

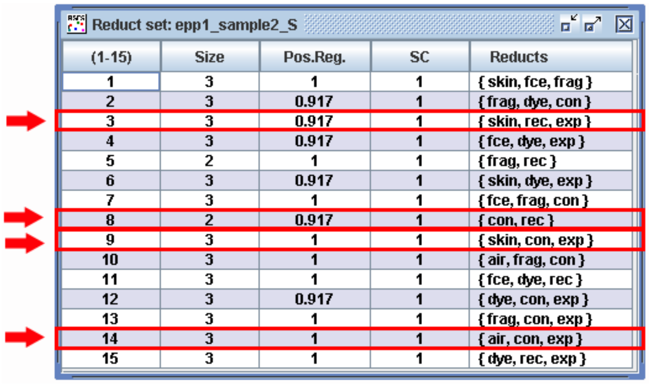
Table 1. Features and their associated cluster centres, in parenthesis, as per each user cluster. Abbreviations are defined in the introduction.

The result of the reduct filtering process is illustrated in Figure 3. The reducts marked by an arrow and encapsulated in a bolded box represent reducts chosen for further analysis. We observe that there are 4, out of the 15 original reducts, that qualify as candidates for further observation. We also observe that of the 4 reducts, the number of features in each ranges between 2 and 3 attributes. Thus, less than 50% of the total system attributes are required to discern the system users based on their feature preferences.

Continuing our algorithm, we test each remaining reduct, one at a time, by clustering the users in the test set using the k-means clustering algorithm, with the same k-value as the training set (k = 2), and the features within the current reduct being observed. The result of the clustering is tested using the cross validation procedure in RSES. Optimal reducts are chosen by first observing those with the least number of dimensions (features). As well, reducts are chosen by those

[5]Version 2.2, freely available online: `http://logic.mimuw.edu.pl/~rses/` (Accessed July 2005).

that, when tested using the cross validation method, have an optimal percentage of correctly classified cases. The results of the cross validation method and a discussion on the success of our algorithm is discussed in the next section.



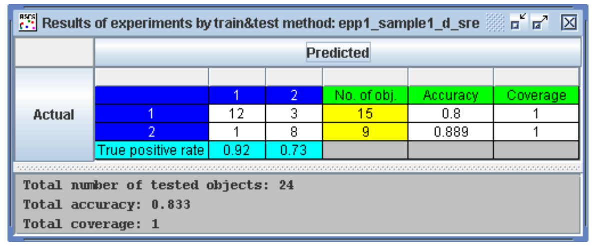
Reduct set: epp1_sample2_S

(1-15)	Size	Pos.Reg.	SC	Reducts
1	3	1	1	{ skin, fce, frag }
2	3	0.917	1	{ frag, dye, con }
3	3	0.917	1	{ skin, rec, exp }
4	3	0.917	1	{ fce, dye, exp }
5	2	1	1	{ frag, rec }
6	3	0.917	1	{ skin, dye, exp }
7	3	1	1	{ fce, frag, con }
8	2	0.917	1	{ con, rec }
9	3	1	1	{ skin, con, exp }
10	3	1	1	{ air, frag, con }
11	3	1	1	{ fce, dye, rec }
12	3	0.917	1	{ dye, con, exp }
13	3	1	1	{ frag, con, exp }
14	3	1	1	{ air, con, exp }
15	3	1	1	{ dye, rec, exp }

Figure 3. The reducts chosen for the train and test procedure as per our filtering algorithm. Those reducts emphasized by arrows and encapsulated in boxes, i.e. reducts 3, 8, 9, and 14, were the ones that passed our filter algorithm.

3 Case Study Results

This section will discuss the results from our case analysis. Of the 15 reducts that were formulated in the training set only 4 of the reducts were selected for further analysis. When testing the reduct {*skin irritation (skin), product uses recyclable packaging (rec), and product minimizes exposure to concentrate (exp)*}, i.e. reduct number 3, we achieve a successful clustering of 83% of the users in the test sample, as illustrated in Figure 4. When testing the reduct {*product is a concentrate (con) and product uses recyclable packaging (rec)*}, i.e. reduct number 8, we achieve a successful clustering of 100% of the users in the test sample, as illustrated in figure 5. At this point we observe that this reduct is the most optimal one out of the four since it has the least number of dimensions and the highest possible accuracy, keeping in mind that the positive region of this reduct is less than optimal at 92%. Because the positive region is not exactly 100% we must also test the remaining 2 reducts, {*skin irritation (skin), product is a concentrate (con), and product reduces exposure to concentrate (exp)*} and {*air pollution potential (air), product is a concentrate (con), and product reduces exposure to concentrate (exp)*}, i.e. reduct numbers 9 and 14, have a successful clustering of 88% and 79% of users in the test set respectively, as illustrated in Figures 6 and 7.

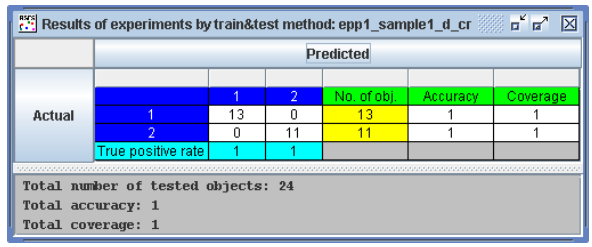


Results of experiments by train&test method: epp1_sample1_d_sre

		Predicted				
Actual		1	2	No. of obj.	Accuracy	Coverage
	1	12	3	15	0.8	1
	2	1	8	9	0.889	1
	True positive rate	0.92	0.73			

Total number of tested objects: 24
Total accuracy: 0.833
Total coverage: 1

Figure 4. Train and test procedure using the reduct set {skin, rec, exp}

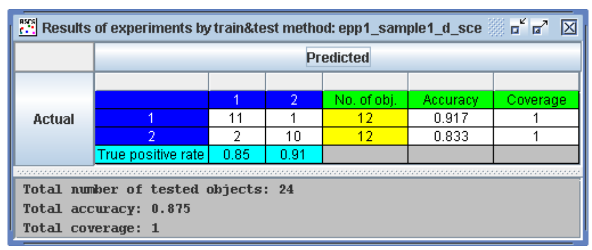


Results of experiments by train&test method: epp1_sample1_d_cr

		Predicted				
Actual		1	2	No. of obj.	Accuracy	Coverage
	1	13	0	13	1	1
	2	0	11	11	1	1
	True positive rate	1	1			

Total number of tested objects: 24
Total accuracy: 1
Total coverage: 1

Figure 5. Train and test procedure using the reduct set {con, rec}



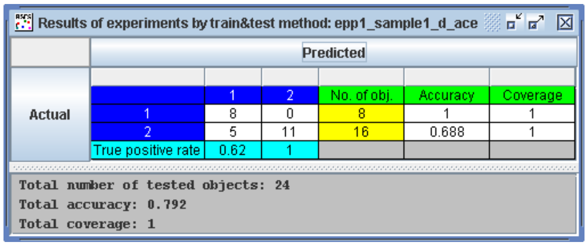
Results of experiments by train&test method: epp1_sample1_d_sce

		Predicted				
Actual		1	2	No. of obj.	Accuracy	Coverage
	1	11	1	12	0.917	1
	2	2	10	12	0.833	1
	True positive rate	0.85	0.91			

Total number of tested objects: 24
Total accuracy: 0.875
Total coverage: 1

Figure 6. Train and test procedure using the reduct set {skin, con, exp}

This verified that the reduct containing the attributes {*product is a concentrate (con) and product uses recyclable packaging (rec)*} proved the most optimal reduct to use when classifying future system users. When observing the positive region in the training set and the percentage of correctly classified users in the test set we can see the potential success of our algorithm. When constructing personalized interfaces for EDSS, or any decision support system (DSS) for that matter, it seems highly probable that by asking users their preference of successful reduct attribute candidates, as per our algorithm, instead of asking users their preference of all features, that we can construct personalized EDSS user interfaces simply by using this information.


Results of experiments by train&test method: epp1_sample1_d_ace

Actual \ Predicted	1	2	No. of obj.	Accuracy	Coverage
1	8	0	8	1	1
2	5	11	16	0.688	1
True positive rate	0.62	1			

Total number of tested objects: 24
Total accuracy: 0.792
Total coverage: 1

Figure 7. Train and test procedure using the reduct set {air, con, exp}

4 Conclusion

The state of our natural environment has become key concern of late. Individual consumers have been asked to take the "challenge" and consider the impact their spending habits and everyday activities produce. However, consumers have had little support concerning ways to be more eco-effective in their everyday lives. There exists EDSS, like the one analyzed in this paper, that allow users to compare alternatives. However, these systems do little to enhance the user's experience while interacting with the system. Personalized interfaces, for EDSS as well as other DSS, is an important area with many interesting problems. Personalization would enhance the user's experience while interacting with the system and thus make their exploration more satisfying.

One way to construct personalized interfaces for EDSS is to ask users to rank the system features according to their perceived preferences. However, as the number of system users increases and if the number of system features is large, this task becomes quite difficult. Clustering users in terms of the feature rankings so as to discover similar groups of users would help reduce the number of interfaces that need to be constructed. As well, clustering users by only asking them to rank those system features necessary to discern clusters would greatly reduce the user initialization process and limit user information overload. The primary goal of our analysis was to see the potential to personalize EDSS user interfaces by clustering users based on their ranked attribute preferences. An underlying goal of simplifying the clustering technique and shortening the user initialization process, i.e. the initial user clustering, was also sought. Our analysis illuminated the potential of our derived algorithm to achieve these goals.

Future work will include the implementation of an EDSS that supports construction of personalized interfaces using the techniques described within this paper. As well, a complimenting case study, similar to the one described in this paper, of the system will be performed and results analyzed.

Acknowledgment

This work was supported by the Centre for Sustainable Communities at the University of Regina, Regina, Saskatchewan. This work was also supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

[1] S. Abidi, K. Meng, and A. Goh. Analyzing data clusters: A rough set approach to extract cluster-defining symbolic rules. *Lecture Notes in Computer Science, Springer Verlag*, 2189, 2001.

[2] J. Bazan, A. Skowron, D. Slezak, and J. Wroblewski. Searching for the complex decision reducts: The case study of the survival analysis. *International Symposium on Methodologies for Intelligent Systems*, pages 160–168, 2003.

[3] J. Bazan and M. Szczuka. The rough set exploration system. *Transactions on Rough Sets III, LNCS 34400, Springer Verlag*, pages 37–56, 2005.

[4] P. Dayan. Unsupervised learning. *The MIT Encyclopedia of the Cognitive Sciences*, 1999.

[5] J. G. Dy and C. E. Brodley. Feature selection for unsupervised learning. *Journal of Machine Learning Research*, 5, 2004.

[6] N. Eagle. Feature selection analysis. 2002.

[7] I. Guyon and A. Elissee. An introduction to variable and feature selection. 2003.

[8] M. H. C. Law, M. A. T. Figueiredo, and A. K. Jain. Simultaneous feature selection and clustering using mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004.

[9] W. McDonough and M. Braungart. Five steps towards reinventing the world, 2001. Online: `http://www.mcdonough.com/writings/five_steps_5.htm` (Accessed June 2005).

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

[11] Z. Pawlak, J. Grzymala-Busse, R. Slowinski, and W. Ziarko. Rough sets. *Communications of the ACM*, 38(11), 1995.

[12] T. Rashid. Clustering, 2003. Online: `http://www.cs.bris.ac.uk/home/tr1690/documentation/fuzzy_clustering_initial_report/node11.html` (Accessed July 2005).

[13] R. R. Tan. Rule-based life cycle impact assessment using modified rough set induction methodology. *Environmental Modeling and Software, Elsevier Science*, 20, 2005.

[14] C. F. Thapalia. Multivariate statistics: An introduction. The Web Center for Social Research Methods, Online: `http://www.socialresearchmethods.net/tutorial/Flynn/multivar.htm` (Accessed July 2005).