ANALYSIS OF USER CLASSIFIERS FOR PERSONALIZATION OF ENVIRONMENTAL DECISION SUPPORT SYSTEM INTERFACES

TIMOTHY MACIAG AND DARYL H. HEPTING
Department of Computer Science, University of Regina
3737 Wascana Parkway, Regina, Saskatchewan,
S4S 0A2, Canada
{maciagt, hepting}@cs.uregina.ca

ABSTRACT
Consumers in current society have become increasingly concerned with the negative impacts their products purchased have on both their individual health and the natural environment. However, consumers have received very little support in their attempt to adopt more eco-effective purchasing habits. There exist environmental decision support systems (EDSS) that allow consumers compare alternatives. However, current EDSS do little to enhance the consumers experience while interacting with the system thus limiting the positive impacts possible. Correct classification of users in terms of their *values* may help personalize each user's view thus make their experience of searching product information more satisfying. This paper discusses common user classification techniques that could be used in classifying EDSS users with the purpose of constructing personalized EDSS user interfaces. A deeper analysis of a classification technique implemented using rough set theory is discussed.

INTRODUCTION

In current societal thought there is an elevated yearning to achieve sustainable practices in our everyday decision-making (OECD, 2002). However, in our attempts to formulate more eco-effective[1] decisions when purchasing commonly consumed products, e.g. household cleaning products, we are often left on our own to decide which products are *best* for us. Knowing what features we require in our product choices, e.g. recyclable packaging or price, helps us formulate these decisions. However, alternatives with fewer impacts, such as cradle-to-cradle products[2] (McDonough and Braungart, 2002), that may exist and be overlooked given the obscurity of relevant product information. It is clear that an environmental decision support system (EDSS) that allows system users compare alternatives would greatly increase adoption of more sustainable and eco-effective purchasing habits. However, current EDSS do little to enhance the user's experience while interacting with the system thus limiting the positive impacts possible.

---

[1] The term eco-effective refers to the concept of producing and consuming products that have a positive and lasting effect on our health and the environment, e.g. a 100% biodegradable fabric, in comparison to eco-efficient products which have a positive, but limiting effect on our health and the environment, e.g. a recyclable plastic bottle (McDonough and Braungart, 2002).

[2] Cradle-to-cradle products refer to products that are eco-effective. Types of these products have a positive and lasting impact on our health and natural environment (see footnote 1 for an example).

To better understand the functionality and usability of current EDSS we performed an analysis of an EDSS based on the United States Environmental Protection Agency's (US-EPA) Environmentally Preferable Purchasing (EPP) Wizard[3]. The results of our analysis have shown that users would greatly benefit from personalization of the EDSS user interface. Personalization of EDSS user interfaces would help individualize each user's view thus making their experience of searching relevant product information more satisfying. Classifying system users based on their perceived *values* of the system features is one approach to acquire the appropriate information to construct personalized EDSS user interfaces. Those system users who perceive features similarly could be grouped where a user interface could be constructed to highlight those features common to the group's values.

There exist a variety of decision-making tools to assist in user classification problems. These can include the multiple attribute utility theory (MAUT) and the analytic hierarchy process (AHP) (Schmitt et al, 2002). These techniques are largely based on ranking/weighting and comparing system features. The ranking of all system features is required when using such methods. However, in many cases, as well as in natural human decision-making, not all features are required to make the actual decision (Pawlak, 1991). Rough set theory, introduced by Pawlak in the early 1980's (Pawlak et al, 1995), provides useful techniques to reduce redundant information thus individual users need only rank/weight and compare those system features required to discern them from other users. This paper provides a brief analysis of common decision-making tools that can be used for classification. As well, a deeper analysis on the use of rough set theory for classification is performed.

The rest of this paper is organized as follows. The next section will discuss common decision-making tools we can use for user classification, specifically MAUT and AHP. The following section will provide an introduction to rough set theory and discuss how it can be used for user classification. The analysis and procedure section describes the algorithms formulated and procedures used in our analysis of a user classification technique that utilized elements of rough set theory. The results section will discuss the results we acquired from our analysis. Finally, the last section describes our conclusions and provides a discussion of relevant future research.

COMMON DECISION-MAKING TOOLS

There are many common decision-making tools we can use to assist in user classification problems. These could include MAUT and AHP as well as others, such as the outranking methods ELECTRE and PROMETHEE where outranking indicates a dominance of one alternative over another (Tan, 2005). This section provides a brief overview of these common classification tools, specifically MAUT and AHP, and how they can assist in classifying users with the goal of constructing personalized EDSS user interfaces.

The multiple attribute utility theory (MAUT) is a decision-making tool that enables the evaluation and/or comparison of objects (Schaefer, 2001). MAUT

---

[3] United States Environmental Protection Agency's Environmentally Preferable Purchasing wizard that compares cleaning products, online: http://www.epa.gov/opptintr/epp/cleaners/select/matrix.htm partially functional (Accessed April 2005).

enables the formulation of decisions based on a decision-maker's pre-defined perception of the strengths and weaknesses of the system features in comparison to each other (Schaefer, 2001). The decision-maker, or in the case of personalizing user interfaces for EDSS, the system users, assign weights to the system features according to their perceived importance. The weighted features are than evaluated using an additive function, referred to as the *utility function,* which amasses a decision-maker's feature preferences (Kabassi and Virvou, 2003).

When considering the issue of constructing personalized user interfaces for EDSS, system users would weight all system features and evaluate them using their pre-defined utility function. Classification of users who weigh features and define their utility function similarly could be grouped together where a personalized user interface could be constructed to highlight the groups' specific feature criteria.

Alternatively, AHP, introduced by Saaty in the early 1980's (Matsuo and Ito, 2003), is a technique similar to MAUT in terms that it is based on system feature rankings. However, rather than evaluating features using a utility function, AHP performs pairwise comparisons of the system features. In many cases, the pairwise feature comparison is performed using a 9-point scale. For example, the following 9-point scale has been commonly used:

- *Features are equally important* – assign a value of 1
- *One feature is slightly more important* – assign a value of 3
- *One feature is strongly more important* – assign a value of 5
- *One feature is very strongly more important* – assign a value of 7
- *One feature is absolutely more important* – assign a value of 9.

Consumers with similar pairwise feature preferences could be grouped where a personalized user interface could be constructed to highlight the groups' specific feature criteria.

Both MAUT and AHP provide useful techniques to accumulate feature preferences and classify system users accordingly. However, both techniques suffer from the requirement that all system features need be weighted/ranked. As an alternative classification technique we can consider utilizing those offered in rough set theory. Rough set theory provides techniques for reducing redundant information in data sets. Thus, system users need rank only those system features necessary to discern them from other system users. The next section will introduce various aspects of rough set theory.

ROUGH SETS

Rough set theory provides techniques for representing uncertainty in knowledge systems (Tan, 2005) and enables the conceptualization of approximations based on data classifications or decision variables (Abidi et al, 2001). The techniques of rough sets as used in our analysis pertain to analyzing the potential of using the minimal set(s) of condition attributes, i.e. system features, necessary to distinguish system objects. These set(s) of features are referred to as *reduct(s).* In order to understand how we formulate reducts we must first understand the basis of rough set theory.

The basis of rough set theory lies in the construction of *information systems*, also commonly known as *decision systems* (Komorowski et al, 1998). The most basic rough set techniques include formulating the *lower* and *upper approximations* as well as the *boundary region* of the equivalence relations in a decision system (Pawlak et al, 1995). For instance, let us consider the following decision system illustrated in Table 1:

Table 1. An example of a decision system and associated approximations.

- Decision System $D = (U, R)$, where:
  - U is the universe of all objects in the decision system,
  - R is an equivalence relation (class), where $R \subseteq D$,
  - $X \subseteq U$, where X is an object(s) of U.

  - Lower Bound = LOW(X) = $\{x \mid [x]_R \subseteq X\}$
  - Upper Bound = UPP(X) = $\{x \mid [x]_R \cap X \neq \emptyset\}$
  - Boundary Region = BND(X) = UPP(X) - LOW(X)

Generally speaking, the lower approximation on a set of data encapsulates all objects belonging without question to a specific classification. The upper approximation on a set of data encapsulates all objects that *may* or *may not* belong to a specific classification. The boundary region on a set of data refers to those objects that cannot be classified with certainty as belonging to a specific classification.

To better understand these relationships we can visualize a decision system in the form of a decision table, i.e. a tabular visualization with rows and columns representing the appropriate system data (Pawlak, 1991). For instance, let us consider the following decision table illustrated in Table 2:

Table 2. A decision table illustrating possible EDSS data. Here, as in our analysis, system users are stating which condition attributes (system features) that they perceive as *very important* by indicating a "yes" value. The system feature "Air Pollution" is randomly chosen as the decision class.

- Decision Table $T = (U, R, C, D)$, where:
  - U and R are as described above,
  - $X \subseteq U$ as described above,
  - C is a condition attribute (system feature),
  - D is a decision class.

X **User**	***C1*** **Fragrance**	***C2*** **Dye**	***C3*** **Recyclable Packaging**	***D*** **Air Pollution**
1	Yes	Yes	Yes	Yes
2	No	Yes	No	No
3	Yes	No	Yes	Yes
4	Yes	No	No	No
5	No	Yes	No	Yes

If we let R be induced by all condition attributes, i.e. $R = \{C1, C2, C3\}$, and consider all negative cases we achieve the following:

- LOW(X) = {User(s): 4},
- UPP(X) = {User(s): 2, 4},
- BND(X) = {User(s): 2},

Another basis of rough set theory includes reduct and *core* formulation. We have already mentioned the concept of reducts. A core, similarly to a reduct, is part of the set of condition attributes necessary to distinguish system objects. However, the main distinction of a core is that it is a condition attribute that is present in every reduct.

One approach for formulating reducts (and cores) is the construction of a *discernibility matrix* (Pawlak, 1991). The main idea of a discernibility matrix is to keep a record of the condition attributes for each object that differ from other objects belonging to different decision classes. The formulation of a discernibility matrix is an iterative process. When completed, the reduct(s) and core(s) will be visualized. Figure 1 illustrates the discernibility matrix based on the decision table depicted in Table 2.

	1	**2**	**3**	**4**	**5**
1	///	///	///	///	///
2	F, R	///	///	///	///
3	///	F, D, R	///	///	///
4	D, R	///	**R**	///	///
5	///	///	F, D, R	**F, D**	///

Figure 1. A discernibility matrix based on the decision table seen in Table 2. Note that the system features are abbreviated as follows: Fragrance (F), Dye (D), and Recyclable Packaging (R). Objects are represented in the first column and row respectively. Objects are compared and records of the condition attributes that differ as per the decision class are visualized in the cells. A "///" signifies that no condition attributes differ when comparing objects OR that the objects belong to the same decision class. The bolded condition attributes represent reduct attribute candidates (as described below).

The first step after the construction of the discernibility matrix is to observe cells containing only one condition attribute. These condition attributes are the first to be recorded as reduct attribute candidates. As seen in Fig. 1, the condition attribute "recyclable packaging (R)" applies. Cells containing the condition attributes in the previous step can be omitted from further observation.

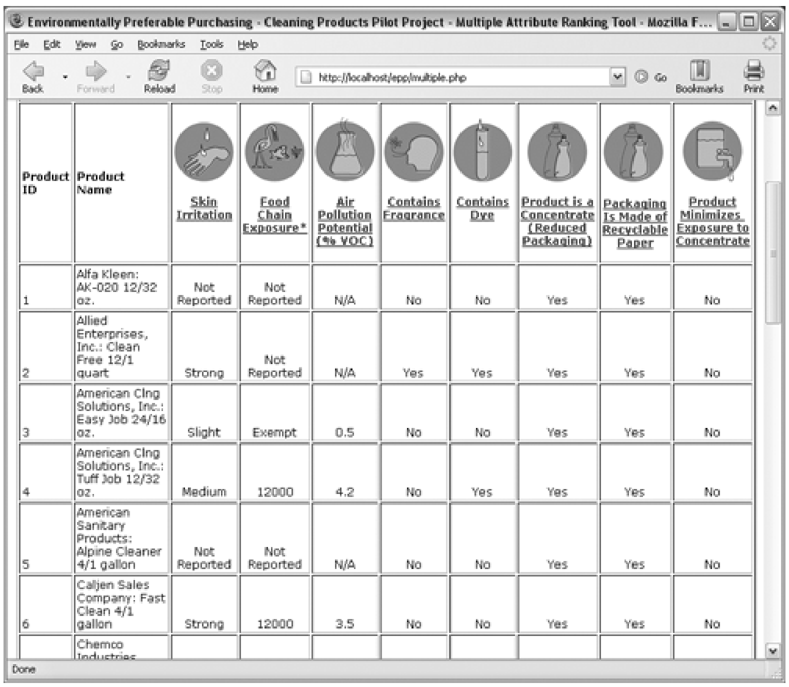
The next step is to apply an additive approach where each of the condition attributes remaining in the cells are accumulated and recorded. The condition attribute with the highest number of cell representations is considered the next reduct attribute candidate. This iterative process is continued until all condition attributes as represented in the cells are accounted for. In the example illustrated in Table 2 and Fig. 1 respectively, the reduct(s) and core(s) would include:

- Reduct 1: {*Recyclable Packaging (R)*, *Fragrance (F)*}.
- Reduct 2: {*Recyclable Packaging (R)*, *Dye (D)*}.
- Core: {*Recyclable Packaging (R)*}.

As illustrated in the above example it becomes apparent that not all condition attributes need to be used to discern objects in this particular example. The primary goals of our analysis included classification of EDSS users based on the system feature preferences. The purpose of this classification was to identify the potential of personalizing EDSS user interfaces accordingly with an underlying goal of simplifying the classification process. Rough set theory provides the techniques to satisfy these goals. The next section will discuss our analysis, which used rough set theory to perform classification of EDSS users based on their system feature preferences, in detail.

ANALYSIS AND PROCEDURES

To test the applicability of rough set theory techniques to perform accurate classification of EDSS users based on their system feature preferences we performed a case study of an EDSS that was based on the United States Environment Protection Agency (US-EPA) Environmentally Preferable Purchasing (EPP) wizard that compares cleaning products, as illustrated in Fig. 2.



Product ID	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
1	Alfa Kleen: AK-020 12/32 oz.	Not Reported	Not Reported	N/A	No	No	Yes	Yes	No
2	Allied Enterprises, Inc.: Clean Free 12/1 quart	Strong	Not Reported	N/A	Yes	Yes	Yes	Yes	No
3	American Clng Solutions, Inc.: Easy Job 24/16 oz.	Slight	Exempt	0.5	No	No	Yes	Yes	No
4	American Clng Solutions, Inc.: Tuff Job 12/32 oz.	Medium	12000	4.2	No	Yes	Yes	Yes	No
5	American Sanitary Products: Alpine Cleaner 4/1 gallon	Not Reported	Not Reported	N/A	No	No	Yes	Yes	No
6	Caljen Sales Company: Fast Clean 4/1 gallon	Strong	12000	3.5	No	No	Yes	Yes	No
	Chemco								

Figure 2. A screen capture of the US-EPA EPP wizard user interface. Cleaning products are visualized in tabular form, with the cleaning products represented in the rows and system features represented in the columns.

48 participants were gathered from the University of Regina Computer Science Participant Pool Program[4]. As well as performing tasks on the system to evaluate the usability of the EDSS user interface, participants were asked to rank the system features according to their perceived importance. Of the 48 participants, 46 actually gave rankings thus only 46 participants were used in our calculations. System features were ranked using a 4-point ranking scale, i.e. *very important*, *important*, *somewhat important* and *unimportant*. The system features ranked included:

1. ***Skin irritation (skin)***, meaning the presence of chemicals in the cleaning product that cause redness or swelling of skin.
2. ***Food chain exposure (fce)***, meaning 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.
3. ***Air pollution potential (air)***, meaning products that may contain volatile organic compounds (VOC), i.e. compounds that have the potential to form atmospheric pollutants, e.g. smog.
4. ***Product contains fragrance (frag)***, meaning fragrances that are added to the cleaning product to improve, or mask, its “natural” odor.
5. ***Product contains dye (dye)***, meaning dyes that have been added to the cleaning product to change the color of the product.
6. ***Product is a concentrate (con)***, meaning cleaning products that are packaged using reduced packaging.
7. ***Product uses recyclable packaging (rec)***, meaning cleaning products that are packaged using recyclable packaging.
8. ***Product minimizes exposure to concentrate (exp)***, meaning cleaning products that reduce exposure to concentrated packaging.

We accumulated the participant system features rankings and recoded them to distinguish between those features participants ranked *very important* and all others. For our analysis we wanted to find system features that would be good candidates for possible decision classes since one of our goals was to cluster system users based on their features preferences. Therefore, we filtered the system features further to distinguish those features which participants had very strong opinions about, **either way**. Meaning, we chose to explore the features that participants **overwhelmingly** ranked as either *very important*, or “*other”* based on the recoded system feature rankings. The result of this filtering process is illustrated in Table 3.

---

[4] The University of Regina Computer Science Participant Pool Program is a program designed to assist researchers in acquiring data for their research projects. Undergraduate students enrolled in eligible computer science and psychology courses are able to enroll in the program where they are asked to participate in research activities at the university. As reward for their participation, students receive a course credit (1%) in their eligible course. Students can receive up to a 2% credit in each eligible course for each time they participate in a research activity.

Table 3. Accumulation of participant rankings of the system features and indication of further observation. "% ranked as {very important, other}" represents the accumulated total percentage of the system feature that participants ranked as very important or gave another ranking.

System Feature	% ranked as very important	% ranked as *other*	Chosen for Further Observation
air	56%	44%	No
skin	54%	46%	No
con	28%	72%	Possibly
rec	28%	72%	Possibly
fce	24%	76%	Possibly
frag	17%	83%	Yes
exp	13%	87%	Yes
dye	9%	91%	Yes

Based on the results illustrated in Table 3, we see that participants had very strong opinions about the system features: *product contains dye (dye), product minimizes exposure to concentrate (exp),* and *product contains fragrance (frag)*. As well, we see that participants had moderate opinions about the system features: *food chain exposure (fce)*, *product uses recyclable packaging (rec)*, and *product is a concentrate (con)*. Thus, these 6 system features were selected as candidates for decision classes and selected for further observation.

For our analysis we utilized the Rough Set Exploration System (RSES)[5]. RSES provides many of the basic rough set functions that were necessary for our analysis. We constructed a decision table consisting of all 46 participants, i.e. the objects, and their associated rankings of the 8 system features, i.e. the condition attributes and decision classes. Each of the 6 system features that were selected for further observation was tested as a decision class variable.

To test the success of the classification of each of the decision class variable candidates, we split the decision table into 2 equally distributed samples to represent training and testing samples respectively. Each sample was randomly chosen as being either the training or testing sample in each case. Using the training sample(s) we formulated the reducts (if any). If more than one reduct was found we constructed sub-table(s) containing only those system features in each of the reduct(s). This was done to test each reduct separately to attain information on the strength of the system features in each reduct to successfully classify potential future system users, i.e. those users in the testing sample. Finally, we tested the classification on the testing sample(s) and observed the number of successfully classified system users.

RESULTS

The results from our analysis were encouraging. As illustrated in Table 4, of the 6 system features selected as decision class candidates, 3 were observed to have a successful classification of participants in the testing samples of 80% or higher. By observing the results in table 4 we can acknowledge that when the system feature, *product contains dye (dye)* is assigned as the decision class

[5] RSES freely available online at: http://logic.mimuw.edu.pl/~rses/ (Accessed May 2005).

variable we can reduce the dimensionality by greater than 50%, i.e. only 3 of the 7 remaining system features (minus the decision class variable) were required to discern system users and still achieve a successful classification of "future system users", as depicted by those users in the testing sample, of 91%.

Table 4. Results from our analysis. Accuracy refers to correctly classified objects in the testing set based on results in the training set.

Decision Class	Reduct(s)	Accuracy
dye	{skin, fce, frag}	91%
exp	{skin, fce, air, con, rec}	80%
frag	{skin, fce, air, con, rec}	76%
fce	{exp, air, frag, con, rec} {skin, exp, air, dye, con, rec}	60% 63%
rec	{fce, air, con, exp} {skin, fce, frag, con, exp}	62% 81%
con	{skin, frag, exp, rec}	75%

The instability of decision classes where accuracies were less than 80% are concerning. However, this instability may indicate that there are strong inconsistencies between the decision class candidate and the other system features based on how participants ranked them. For instance, when the feature *product contains dye* was assigned as the decision class variable, users had little interest in this feature thus one decision class, i.e. the users who did not rank *product contains dye* as *very important*, were overwhelmingly represented. Perhaps more information about system users is required before classification and feature reduction can be performed, i.e. what exactly are the *values* of the system users in relation to all system attributes, etc. However, given the results of our analysis, the potential to efficiently classify EDSS users with the goals of constructing personalized EDSS user interfaces based on system user feature rankings using the described approach appears promising.

CONCLUSIONS AND FUTURE WORK

This paper performed an analysis of various classification techniques with an emphasis on rough set techniques for classification. The primary goals of our analysis was to classify EDSS users based on their system feature preferences to identify the potential to construct personalized EDSS user interfaces using the classification information with an underlying goal of simplifying the classification process.

A brief analysis of common decision-making tools that could be used for user classification, specifically MAUT and AHP, was performed with a deeper analysis of rough set classification techniques. Although it was observed that the common techniques prove useful in classification problems, given that users need to classify all system features before classification can proceed, an alternate approach can be considered. Rough set theory provides techniques to remove redundant information in decision systems. Thus, users need only rank those system features necessary to discern themselves from others. Based on the results of our analysis, the promise of using rough sets for classification of EDSS users appears quite promising.

Future work will include a deeper analysis of the inconsistencies seen in the results of our analysis. A modified technique is currently being developed that uses elements of multivariate statistics, rough sets, and machine learning to classify EDSS users with the same primary goals as those described in this paper. Specifically, the k-means clustering algorithm is being used to formulate user clusters wherein the user cluster value essentially becomes the decision class variable. Rough set techniques are than performed to reduce the dimensionality of the system features. The algorithm is being tested using a train/test procedure similar to that described in this paper.

ACKNOWLEDGEMENTS

We would like to acknowledge Dr. Dominik Slezak of the University of Regina for his contributions in refining our concept and approach. This work was supported by the Centre for Sustainable Communities at the University of Regina and by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Abidi, S, Meng, K, and Goh, A. "Analyzing data clusters: A rough set approach to extract cluster-defining symbolic rules". Lecture Notes in Computer Science, (2189), 2001.

Organization for Economic Co-operation and Development (OECD). "Report of the OECD workshop on information and consumer decision-making for sustainable consumption", 2002.

Kabassi, K., and Virvou, M. "Combination of a cognitive theory with the multi-attribute utility theory". Lecture Notes in Computer Science, 2773:944–950, 2003.

Komorowski, J., Pawlak, Z., Polkowski, L., and Skowron, A. "Rough sets: A tutorial". Rough-Fuzzy Hybridization: A New Method for Decision Making, 1998.

Linkov, I., Varghese, A., Jamil, S., Seager, T. P., Kiker, G., and Bridges, T. "Multi-criteria decision analysis: A framework for structuring remedial decisions at contaminated sites". Comparative Risk Assessment and Environmental Decision Making, pages 15–54, 2004.

Matsuo, T. and Ito, T. "A decision support system for group buying based on buyer's preferences in electronic commerce". The 20th International World Wide Web Conference, 2003.

McDounough, W. and Braungart, M. "Design for the triple top line: New tools for sustainable Commerce". Corporate Environmental Strategy, 9(3), 2002.

Pawlak, Z. "Rough Sets, Theoretical Aspects of Reasoning about Data". Kluwer Academic Publishers, 1991.

Pawlak, Z, Grzymala-Busse, J, Slowinski, R, and Ziarko, W. "Rough sets". Communications of the ACM, 38(11):88–95, 1995.

Schaefer, R. "Rules for Using Multi-Attribute Utility Theory for Estimating a User's Interest". Agent Based Information Systems (ABIS) Workshop, 2001

Schmitt, C., Dengler, D., and Bauer, M. "The maut machine: An adaptive recommender system". Agent Based Information Systems (ABIS): Personalization for the Mobile World, 2002.

Tan, R. R. Rule-based life cycle impact assessment using modified rough set induction methodology". Environmental Modeling and Software, (20), 2005.