Ranking by Rough Approximation of Preferences for Decision Engineering Applications

Kazimierz Zaras[1] and Jules Thibault[2]

[1] Département des Sciences de la Gestion
Université du Québec en Abitibi-Témiscamingue Rouyn-Noranda
Québec, Canada J9X 5E4
Kazimierz.Zaras@uqat.ca
[2] Department of Chemical Engineering, University of Ottawa
Ottawa, Ontario, Canada K1N 6N5
Jules.Thibault@uottawa.ca

Abstract. A pulping process is studied to illustrate a new methodology in the field of decision engineering, which relies on the Dominance Rough-Set-based Approach (DRSA) to determine the optimal operating region. The DRSA performs a rough approximation of preferences on a small set of Pareto-optimal experimental points to infer the decision rules with and without considering thresholds of indifference with respect each attribute in the decision table. With thresholds of indifference, each rule can be represented by three discrete values (i.e. 0; 0.5; 1). A value of (1) indicates the first point, in a pair wise comparison, is strictly preferred to the second point from the Pareto domain. A value of (0) indicates the opposite relation whereas a value of (0.5) indicates that the two points are equivalent from an engineering point of view. These decision rules are then applied to the entire set of points representing the Pareto domain. The results show that the rules obtained with the indifference thresholds improve the quality of approximation.

Keywords: Neutral network, Genetic algorithm, Pareto domain, Preferences, Multicriteria analysis, Dominance-based Rough Set Approach.

1 Introduction

During the operation of an industrial process, the operator should ideally select values of input parameters/variables from a performance criterion point of view. The main problem facing the decision maker is that the range of parameter/variable values is usually very large and the number of their combinations is even larger such that a decision aid methodology is required to assist the decision maker in the judicious selection of all values of the process parameter/variables that lead to the best compromise solution in the eyes of the expert that has a profound knowledge of the process. This is at the core of a new decision engineering methodology, which mainly consists of three steps:

1. Process modelling,
2. Determination of the Pareto domain defined in terms of input parameters, and
3. Pareto Set ranking by the Rough Set Method.

J.T. Yao et al. (Eds.): RSKT 2007, LNAI 4481, pp. 142–148, 2007.

The rough set method allows capturing relatively easily valuable, at time unconscious, information, from an expert that a profound knowledge about the operation for the process in other to establish a ranking method that will be used to rank the entire Pareto domain.

2 Process Modelling

This methodology will be illustrated using a pulping process example. In the pulping process it is necessary to choose, using an appropriate multicriteria methodology, the set of operating conditions that will give the optimal quality of the pulp and the resulting paper sheet. The pulping process is a very complex nonlinear process for which a model is not readily available. It is therefore desired to have a model that can predict the various quality characteristics of the final product. To derive this model, a series of experiments were conducted by Lanouette et al. [4] in a pilot-scale pulp processing plant located in the Pulp and Paper Research Centre at Université du Québec à Trois-Rivières.

Among the numerous performance criteria, four objective criteria were retained as the most important ones for this process (see Thibault et al. [7] and Renaud et al. [6] for a more complete description of the process). The aim in this process is to maximize both the ISO brightness (Y_1) and the rupture length (Y_4) of the resulting paper sheet, while reducing the specific refining energy requirement (Y_2) and the extractive contents (Y_3). The experimental design that was used to perform the experiment is a D-Optimal design where seven input variables were considered. A D-Optimal design consists of a group of design points chosen to maximize the determinant of the Fisher information matrix (X'X). To model each of the four performance criteria of the process, stacked feedforward neural networks were used. Each neural network used the seven input process variables.

3 Determination of the Pareto Domain

The next step of the methodology consists of determining the region circumscribing all feasible solutions of the input variables represented by a large number of data points. An extension of the traditional genetic algorithm is suggested to deal with discretized data by introducing the dominance concept (see [3]). The procedure to obtain a good approximation of the Pareto domain is relatively simple. The n points randomly chosen initialize the search algorithm. For each point, the performance criteria are evaluated. A pair wise comparison of all points approximating the Pareto domain is performed. Then a dominance function, consisting of counting the number of times a given point is dominated by the other points, is calculated. A fraction of the dominated points corresponding to those most dominated is discarded. The non-dominated and the least dominated points are retained and recombined to replace the most dominated ones that were discarded. The recombination procedure is applied until all points are non-dominated. In the case of the pulping process, the Pareto domain defined in terms of input variables is represented by 6000 points. This number of points is too numerous to allow the decision-maker to easily select the zone of optimal conditions. For this reason it is necessary to use a ranking algorithm to establish the optimal region of operation. The next step of this overall methodology deals with this problem. The particular method used in this investigation is the Rough Set Method which is based on the Dominance Rough-Set-Based Approach (DRSA) (see [1]).

4 Ranking the Entire Pareto Set Using the Rough Set Method

The Rough Set Method is used to rank a large number of non-dominated points approximating the Pareto domain. The implementation of this ranking scheme is based on the Rough Set theory suggested by Pawlak [5], and developed by Greco et al. [1-2,7] and Zaras [9], a method known as the Dominance Rough-Set–based Approach (DRSA).

The procedure of this ranking method can be summarized as follows (see Thibault et al. [8]). First, a handful of points, usually (4-7), from different regions of the Pareto domain are selected and presented to a human expert who has a profound knowledge of the process. The expert is given the task of ordering the subset of points from the most preferred to the least preferred (Table 1). After creating the ranked subset, the expert specifies the indifference threshold for each criterion. The indifference threshold corresponds to measurement error as well as possible limits in the human detection of differences in a given criterion. Especially, the indifference threshold for a particular criterion is defined as the difference between two values of that criterion that is not considered significant enough to rank one value as preferred over another (Table 2).

Table 1. Subset of points from the Pareto domain ranked by the expert

Point	Y_1	Y_2	Y_3	Y_4
16	66.95	7.04	0.311	4.07
271	69.49	7.64	0.218	3.76
223	68.82	7.26	0.166	3.54
4671	69.46	7.91	0.222	3.89
12	66.99	7.25	0.526	4.14
2	68.68	6.29	0.469	2.55
1	67.85	6.53	0.273	1.94

The next step is to establish a set of rules that are based on the expert's ranked subset and indifference thresholds. Here, each point in the ranked subset is compared to every other point within that set in order to define "rules of preference" and "rules of non-preference". Each rule can be represented by a vector containing two (i.e. 0; 1, see Table 3) or three values (i.e. 0; 0,5; 1, see Table 4) depending if the comparison is performed without or with indifference thresholds. The dimension of each vector is equal to the number of attributes. In the case without thresholds a value of (1) for a given criterion indicates the first point of the compared pair is preferred to the second point whereas a value of (0) indicates the criterion of the second point is preferred to the first point. However, it is not known if the point is weakly or strongly preferred for that particular criterion. In the case with thresholds of indifference, a value of (1) indicates the first point of the compared pair is strictly preferred to the second point from the Pareto domain, a value of (0) indicates the opposite relation and (0.5) indicates the indifference because the gap between the two values of the criterion is not sufficient to allow choosing one point over the other. The conjunctions ($0.5 \vee 1$) and ($0 \vee 0.5$) indicate the weak preference and the weak non-preference, respectively. These decision rules are then applied to the whole set of points approximating the Pareto domain where all pairs of points are compared to determine if they satisfy a preference or a non-preference rule.

Table 2. Indifference thresholds for each criterion

Criterion	Description	Threshold
Y_1	ISO Brightness	0.50
Y_2	Refining Energy	0.40
Y_3	Extractives Content	0.05
Y_4	Rupture Length	0.30

Table 3. Set of rules for the ranked set without indifference thresholds

Preference rules				Non-preference rules			
Y_1	Y_2	Y_3	Y_4	Y_1	Y_2	Y_3	Y_4
0	0	1	1	1	1	0	0
1	1	0	1	0	0	1	0

The quality of the approximation expresses the ratio of all pairs of points in the ranked subset correctly ordered by “rules of preference” and “rules of non-preference” to the number of all the pairs of points in the ranked subset. The quality of approximation in the case without thresholds is equal to 0.38.

The quality of the approximation in the case with thresholds is equal to 0.57, which indicates that the quality of the approximation with the thresholds is significantly improved.

The last step of the Rough Set method is to perform a pair wise comparison of all 6000 points of the Pareto domain to determine if a preference or non-preference rule applies. If a preference rule is determined, the score of the first point is incremented by one and the score of the second point is decreased by one. The opposite operation is performed if a non-preference rule is identified. The scores of all Pareto-optimal points were initially set to zero. When all points are compared, the point that has the highest score is considered to be the optimal point. It is however preferable to examine the zone of the Pareto domain where a given percentage of the best points are located rather than considering an individual point. The Rough Set Method provides a clear recommendation as to the optimal zone of operation.

Table 4. Set of rules for the ranked set with indifference thresholds

Preference rules				Non-preference rules			
Y_1	Y_2	Y_3	Y_4	Y_1	Y_2	Y_3	Y_4
-	-	0.5	1	-	-	0.5	0
-	1	-	$0.5 \vee 1$	-	0	-	$0.5 \vee 0$
0.5	0.5	1	0.5	0.5	0.5	0	0.5
1	0.5	0	0.5	1	0	0	0

5 Results and Conclusions

Results obtained using the Rough Set Method (RSM) are presented in Fig. 1 without thresholds and in Fig. 2 with thresholds. Two-dimensional graphical projections show the results for both cases of the ranking the 6000 points of the Pareto front. The first 10% corresponding to the highly-ranked points are plotted using dark points.

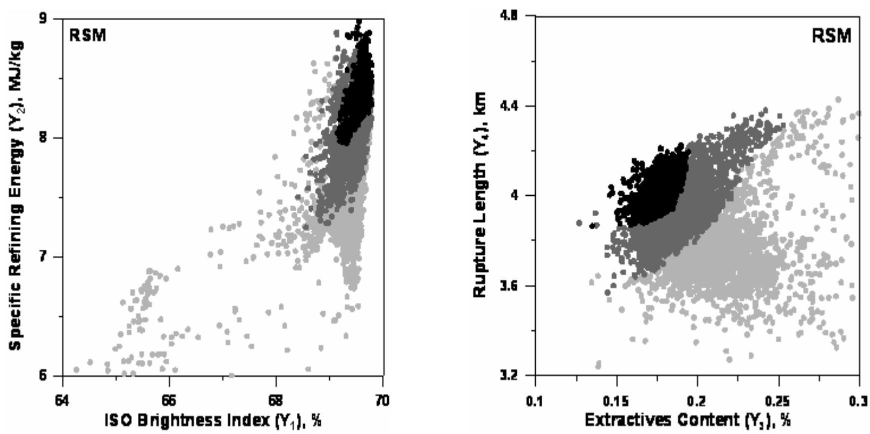

Fig. 1. Graph of the Pareto Front ranked by RSM without thresholds

The optimal region satisfies very well three of the four criteria. The choice of the expert is very clear, he can sacrifice having a higher specific refining energy (Y_2 being highest in the optimal region when it should be lowest) to have all the other criteria (Y_1, Y_3 and Y_4) being satisfied extremely well. In RSM, there is always at least one criterion that has to be sacrificed because a preference rule cannot contain all ones. Indeed, if a rule contained all ones this would mean that one of the two points that led to that rule would dominate the other point.

In this paper, two Dominance Rough–Set-based Approaches have been compared based on the extraction of rules from a subset of Pareto-optimal points ranked by a DM with and without indifference thresholds. The comparison of the quality of approximation indicates that the quality performance is improved with using the thresholds. However, the improved quality doesn't reduce the region of highly-ranked points. On the contrary, we can see on the graphical projections of the Pareto Front that this region is getting larger. There seems to have more nuance in the choice of the decision maker.

The introduction of indifference thresholds to the Dominance Rough-Set-based Approach (DRSA) also allows to make a difference between weaker and strict partial preferences with respect to each criterion for each decision rule.

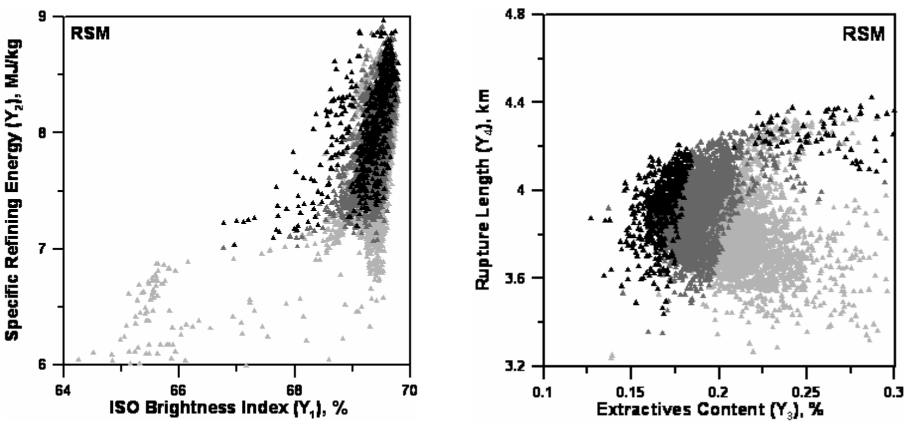

Fig. 2. Graph of the Pareto Front ranked by RSM with thresholds

References

1. Greco, S., Matarazzo B., Slowinski. R.: Rough approximation of a preference relation by dominance relations. EJOR, 117, 2, (1999) 63-83.
2. Greco, S., Matarazzo B., Slowinski. R.: Rough Set methodology for sorting problems in presence of multiple attributes and criteria, EJOR 138, 2, (2002) 247-259.
3. Fonteix, C., Bicking, F. Perrin, E., Marc, I.: Haploïd and diploïd algorithms, a new approach for global optimization: compared performances, Int. J. Sys. Sci. 26, 10, (1995), 1919-1933.
4. Lanouette, R., Thibault, J., Valade, J.L.: Process modeling with neural networks using small experimental datasets, Comp. Chem. Engng, (1999) 23, 1167-1176.
5. Pawlak, Z.: Rough Sets. Theoretical aspects of reasoning about data. Kluwer Academic Publishers, Dordrecht (1991).
6. Renaud, J., Thibault, J., Lanouette, R., Kiss, L.N., Zaras, K., Fonteix., C.: Comparison of two multicriteria decision aid methods: Net Flow and Rough Set Methods in a high yield pulping process, , EJOR 177, 3, (2007) 1418-1432.

7. Slowinski, R., Greco, S., Matarazzo, B.: Rough Set based decision support. Chapter 16 [in] Burke, E.K., and Kendall, G. (eds.), Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques, Springer-Verlag, New York, 2005, pp. 475-527.
8. Thibault, J., Taylor, D., Yanofsky, C., Lanouette, R., Fonteix, C., Zaras, K.: Multicriteria optimization of a high yield pulping process with Rough Sets. Chem. Eng. Sci. 58 (2003) 203-213.
9. Zaras, K.: Rough approximation of preference relation by a multiattribute dominance for deterministic, stochastic and fuzzy decision problems, EJOR, 159, 1, (2004), 196-206.