

# Dominance-based Rough Set Approach for Decision Analysis - a Tutorial

Salvatore Greco  
Faculty of Economics, University of Catania  
Corso Italia 55, 95129 - Catania, Italy  
salgreco@unict.it

Scientific analysis of decision problems aims at giving the decision maker (DM) a recommendation concerning a set of objects (called also alternatives, solutions, acts, actions, cases, candidates) evaluated from the point of view of a plurality of their characteristics considered relevant for the problem at hand, and called *attributes*.

For example, a decision can regard:

- 1) diagnosis of pathologies for a set of patients, being the objects of the decision, and the attributes are symptoms and results of medical examinations,
- 2) assignment to classes of risk for a set of enterprises, being the objects of the decision, and the attributes are ratio indices and other economical indicators such as the market structure, the technology used by the enterprises, the quality of the management and so on,
- 3) selection of a car to be bought from among a given set of cars, being the objects of the decision, and the attributes are maximum speed, acceleration, price, fuel consumption and so on,
- 4) ordering of students applying for a scholarship, being the objects of the decision, and the attributes are scores in different disciplines.

The following three most frequent decision problems are typically distinguished:

- *classification*, when the decision aims at assigning objects to predefined classes,
- *choice*, when the decision aims at selecting the best objects,
- *ranking*, when the decision aims at ordering objects from the best to the worst.

Looking at the above examples, we can say that 1) and 2) are classification problems, 3) is a choice problem and 4) is a ranking problem. The above categorization can be refined with respect to classification problems by distinguishing between

- *nominal classification*, called also *taxonomy*, when neither the value sets of attributes nor the predefined classes are preference ordered,
- *ordinal classification*, called also *sorting*, when both the value sets of attributes and the predefined classes are preference ordered (Greco, Matarazzo and Slowinski 2001) (even if in this case it is also possible to take into account attributes with non ordered value sets (Greco, Matarazzo and Slowinski 2002b)).

Looking at the above examples, one can say that 1) is a nominal classification problem, while 2) is an ordinal classification problem.

If value sets of attributes are ordered according to preference, they are called *criteria*, otherwise they keep the name of attributes. For example, in the decision regarding the selection of a car, its price is a criterion, because, obviously, a low price is better than a high price. Instead, the color of a car is not a criterion but simply an attribute, because red is not intrinsically better than green. One can imagine, however, that also the color of a car could become a criterion, if, for example, a DM would consider red better than green.

For scientific support of decisions one needs a more or less explicit model of the decision problem. The model relates the decision to the characteristics of the objects expressed by the considered attributes. Building such a model requires information about conditions and parameters of aggregation of multi-attribute characteristics of objects. The nature of this information depends on the methodology: prices and interest rates for cost-benefit analysis, cost coefficients in objectives and technological coefficients in constraints for mathematical programming, training sets for neural

networks, substitution rates for value functions of Multi-Attribute Utility Theory, pairwise comparisons of objects in terms of intensity of preference for Analytic Hierarchy Process, attribute weights and several thresholds for ELECTRE methods, and so on. This information has to be provided by the DM, possibly assisted by an analyst.

Very often this information is not easily definable. For example this is the case of the price of many immaterial goods and of the interest rates in cost-benefit analysis, or the case of the coefficients of objectives and constraints in mathematical programming models. Even if the required information is easily definable, like a training set in neural networks, it is often processed in a way that is not clear for the DM, such that he cannot see what are the exact relations between the provided information and the final recommendation. Consequently, very often the decision model is perceived by the DM as a *black box* whose result has to be accepted because the analyst's authority guarantees that the result is "right". In this context, the aspiration of the DM to find good reasons to make decision is frustrated and it rises the need for a more transparent methodology in which the relation between the original information and the final recommendation is clearly shown. Such a transparent methodology searched for can be called *glass box*. Its typical representative is based on application of decision rules which express the decision model in terms of a set of "if..., then..." statements. From one side, the decision rules are explicitly related to the original information and, from the other side, they give understandable justifications for the decision to be made (Greco, Matarazzo and Slowinski 2005).

For example, in case of a medical diagnosis problem, the decision rule approach requires as input information a set of previous diagnoses, from which some rules are induced, such as "if there is symptom  $\alpha$  and the test result is  $\beta$ , then there is pathology  $\gamma$ ". Each one of such rules is directly related to a set of diagnoses in the input information, where there is symptom  $\alpha$ , test result  $\beta$  and also pathology  $\gamma$ . Moreover, the DM can verify by himself that in the input information there is no diagnosis in which there is symptom  $\alpha$  and test result  $\beta$ , but not pathology  $\gamma$ .

The rules induced from the input information provided in terms of exemplary decisions represent a transparent decision model for the DM, and enable his understanding of the reasons of his previous decisions. The acceptance of the rules by the DM justifies, in turn, their use for decision support.

The induction of rules from examples is a typical approach of artificial intelligence. This explains our interest in rough set theory (Pawlak 1991) which proved to be a useful tool for analysis of vague description of decision situations. The aim of rough set analysis is the explanation of the dependence between the values of some decision attributes, playing the role of "dependent variables", by means of the values of other condition attributes, playing the role of "independent variables". For example, in the above diagnostic context, data about the presence of some pathology are given by decision attributes, while data about symptoms and tests are given by condition attributes. An important advantage of the rough set approach is that it can deal with partly inconsistent data, for example, cases where the presence of different pathologies is associated with the presence of the same symptoms and test results. Moreover, it provides useful information about the role of particular attributes and their subsets, and prepares the ground for representation of knowledge hidden in the data by means of "if ..., then ..." decision rules.

Classical Rough Set Approach (CRSA) proposed by Pawlak cannot deal with preference order in the value sets of condition and decision attributes. Thus, from among all listed above decision problems, classical rough set approach supports nominal classification only. For ordinal classification, choice and ranking it is necessary to generalize the classical rough set approach, so as to take into account preference orders. This generalization, called Dominance-based Rough Set Approach (DRSA), has been proposed by Greco, Matarazzo and Slowinski (Greco, Matarazzo, Slowinski 2001 and 2005). Moreover, it has been proved that CRSA is a specific case of DRSA, and, therefore, any application of rough set approach to decision problems can be effectively dealt with using DRSA (Greco, Matarazzo and Slowinski 2007a).

In this tutorial we present applications of DRSA to many real life decision problems, emphasizing the advantages of this approach with respect to competitive approaches. More in detail, we shall consider the following problems.

*Bankruptcy risk evaluation* (Greco, Matarazzo and Slowinski 2002a). This is a typical ordinal classification problem and, therefore, it can be naturally dealt with DRSA; the considered criteria are some financial ratios having a continuous scale such as net income/net worth, total liabilities/total assets, interest expenses/sales and some qualitative ordinal evaluations such as managers work experience, firm's market position, technical structure-facilities and so on. With respect to traditional methodologies, such as discriminant analysis, among many other advantages, DRSA permits to take into account the qualitative ordinal evaluations without any artificial transformation into numerical values.

*Medical diagnosis*. This is a typical nominal classification problem, but after a proper re-codification consisting in a monotonic binarization of nominal attributes, it can be transformed to an ordinal classification problem and, therefore, processed using DRSA; this re-codification permits to refine the modelling of the problem by introducing graduality to important binary attributes, which is not possible within CRSA.

*Budget allocation in highway maintenance activities*. We consider the problem of maintenance funds allocation among various highways owned by a central agency taking into account several criteria. Our methodology allocates the resources on the basis of an ordinal classification of the various highways with respect to the necessity of maintenance. The classification is built using DRSA, because it allows to consider quantitative and qualitative criteria and, starting from some examples of decisions supplied by the DM, gives recommendation in terms of easily understandable "if..., then..." decision rules.

*Evaluation of rural sustainable development potentialities*. Using DRSA, we evaluate a certain number of sites in a region with respect to rural sustainable development potentialities. The proposed evaluation is based on a set of criteria and on a set of examples being reference sites whose rural sustainable development potentialities have been already evaluated by one or more experts. The DRSA is applied to the analysis of the reference sites. It gives a set of decision rules expressing minimal profiles, in terms of considered criteria, for assigning a new site to a given class of sustainable development potentialities.

*Facility services procurement*. We propose a DRSA methodology to support procurement strategies for facilities services in health-care sector. The DRSA has been applied to take into account data affected by inaccuracy and vagueness when even the order of importance of the considered criteria is not clear to the involved manager. The focus is on the comparison of several outsourcing alternatives in facilities management services procurement. The data were collected through interviews with a group of hospital managers.

*Customer satisfaction and Kansei engineering* (Greco, Matarazzo and Slowinski 2007b). DRSA permits to represent the preferences of the customers by means of simple and clearly understandable decision rules, instead of the utility functions considered in usual conjoint measurement analysis. The proposed methodology gives also some indications relative to efficiency of strategic interventions aimed at improving the quality of the offered products and services. The expected efficiency of these interventions is also measurable (Greco, Matarazzo, Slowinski, Pappalardo 2005). This is very useful for the definition and implementation of proper customer satisfaction strategies. The DRSA results are interesting also for Kansei engineering.

*Urban transportation choice*. Application of DRSA to transportation mode choice permits to obtain a set of "if..., then ..." decision rules, which allow to trace out the main components of individual behaviour in transport mode choices. DRSA has been applied to a real data set, referred to a

mobility survey realized in Trieste, a medium town of north of Italy. DRSA results have been compared to those obtained considering Random Utility Model.

*Product mix* (Greco, Matarazzo and Slowinski 2007c). Traditionally this is a mathematical programming problem aiming at defining how much to produce of each product (i.e. the mix), to maximize the profit subject to resource availability constraints. Interactive Multiple Objective Optimization (IMOO) permits to take into account not only maximization of the profit but also other objectives, such as minimization of critical materials or maximization of production of some specific products. DRSA permits to deal with IMOO in a very convenient way. The method is composed of two main stages. In the first stage, a sample of solutions from the Pareto optimal set is generated. In the second stage, the DM indicates relatively good solutions in the generated sample. From this information, a preference model expressed in terms of “if..., then ...” decision rules is induced using DRSA. These rules define some constraints cutting-off non-interesting solutions from the currently considered Pareto optimal set. From the updated Pareto optimal set, a new sample of solutions is generated in the next iteration. The interaction continues until the DM finds a satisfactory solution in the generated sample.

*Financial portfolio selection.* In the context of the classical Markowitz model of financial portfolio, we consider an interactive multiple objective optimization having as criteria to be maximized the gains relative to some meaningful thresholds of the cumulative probability of portfolio returns (for example, the gain one can get with probability of at least 50%, 75% and 95%). From evaluations expressed by the investor on some exemplary portfolios, we induce some decision rules representing the preferences of the investor, using DRSA. These rules are then used to reduce the set of admissible portfolios to be presented to the investor, until he is satisfied and selects his most preferred portfolio.

*Management of environmental and land resource.* The DRSA within the IMOO procedure has been used to define sustainable strategy dealing with the conflict between environmental protection and economic development by different uses of land. The optimization objectives include economic benefits characterized by income and employment level, water quality impacts related to discharges of pollutants under the constraints following from the land-use programs and tolerable water pollution.

### References

- Greco, S., Matarazzo, B., Slowinski, R.: "Rough Sets Theory for Multicriteria Decision Analysis". *European Journal of Operational Research* 129, 1 (2001) 1-47
- Greco, S., Matarazzo, B., Slowinski, R.: "Rough Approximation by Dominance Relations". *International Journal of Intelligent Systems* 17, 1 (2002a) 153-171
- Greco, S., Matarazzo, B., Slowinski, R.: "Rough sets methodology for sorting problems in presence of multiple attributes and criteria". *European Journal of Operational Research* 138, (2002b) 247-259
- Greco, S., Matarazzo, B., Slowinski, R.: "Decision Rule Approach", chapter 13 in J. Figueira, S. Greco, M. Ehrgott (eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer, Berlin, 2005, pp. 507-563
- Greco, S., Matarazzo, B., Slowinski, R.: "Dominance-based Rough Set Approach as a proper way of handling graduality in Rough Set Theory". *Transaction on Rough Sets VII*, LNCS 4400, 36-52, 2007a
- Greco, S., Matarazzo, B., Slowinski, R.: "Customer Satisfaction Analysis based on Rough Set approach". *Zeitschrift für Betriebswirtschaft*, 77, 3 (2007b), 325-339
- Greco, S., Matarazzo, B., Slowinski, R.: "Dominance-based Rough Set Approach to Multiobjective Interactive Optimization", in J.Branke, K.Deb, K.Miettinen, R.Slowinski (eds.), *Multiobjective Optimization: Interactive and Evolutionary Approaches*. Springer-Verlag, to appear, 2007c
- Greco, S., Matarazzo, B., Pappalardo, N., Slowinski, R.: "Measuring Expected Effects of Interventions based on Decision Rules". *Journal of Experimental & Theoretical Artificial Intelligence*, Vol. 17, No. 1-2, January-June 2005, 103-118
- Pawlak, Z.: "Rough Sets. Theoretical Aspects of Reasoning about Data". Kluwer Academic Publishers, Dordrecht, 1991