FINDING COMPACT MODELS USING ROUGH MODELING

Terje Løken and Jan Komorowski
Knowledge Systems Group, Department of Information and Computer Science
Norwegian University of Science and Technology
7491 Trondheim, Norway
Phone: +47 73593440, Fax: +47 73594466
e-mail: {terjelo, janko}@idi.ntnu.no

ABSTRACT: Traditional data mining methods based on rough set theory focus on extracting models which are good at classifying unseen objects. If one wants to discover knowledge, the descriptive qualities of a model are also important, and ideally, one would like to maximize both. Rough Data Modeling is an approach which is suited to this task, in addition to being computationally simple enough to handle large data sets. The theory of rough data modeling has been further investigated, and key concepts have been implemented in a rough set data analysis toolkit, ROSETTA. Initial experiments confirm that the drop in performance of rough models compared to models induced using traditional rough set methods is slight at worst, and the gain in descriptive quality is very large.

KEYWORDS: Knowledge Discovery, Rough Sets, Rough Data Modeling

INTRODUCTION

Rough set theory, introduced by Pawlak (1991), provides a theoretically sound framework for extracting models, in the form of propositional decision rules, from data. There are two main tasks for which a model is useful: *prediction* and *description*. Prediction is concerned with predicting future or unknown values of some attributes using available data, while description means to identify important patterns in the data, and present them to the user in an understandable way.

This means that there are at least two different axes on which to grade a model; predictive and descriptive abilities. When performing knowledge discovery from databases (KDD), we are interested in finding as good a model as possible from a set of data. However, what constitutes a good model may vary, depending on the goals of the particular KDD process. If the goal is to build a model able to classify unseen objects as accurately as possible, the predictive quality is all-important. Several quality measures exist which measure the predictive powers of a model, *accuracy* and *receiver operating characteristics* (ROC) (Hanley and McNeil (1982)) being commonly used. If the goal of the KDD process is description, we need to formalize what it means that a model displays good descriptive qualities. This is a difficult task, but the *size* of the model is of fundamental importance. A model consisting of thousands of rules, or utilizing several hundred attributes is incomprehensible, while a small model, containing in the neighborhood of twenty rules or less, is easily understood by a human.

Very often, we would like to find a model which has good descriptive quality, and still retains good predictive quality. The starting assumption is that a good descriptive model will be a *compact* model, meaning that it contains a limited number of rules, and each rule uses a manageable number of attributes. There are two fundamentally different approaches to finding models displaying these characteristics. The first approach starts out by generating as good a predictive model as possible, ignoring the descriptive quality. It is then believed that within this model, there are some patterns which are of fundamental importance, and others which are redundant or only apply to a small number of objects in the data. Each model consisting of rules which are contained in the original model is called a *submodel*, and using various *filtering* strategies, one accepts a small drop in predictive performance in exchange for a submodel which is significantly smaller than the original model.

The other philosophy, and the one adopted in this paper, is that one should aim to locate the important patterns directly from the data. There are several reasons for doing this. If we use traditional rough set methods to find a good predictive model, this means calculating *reducts*, a problem which is NP-hard (Skowron and Rauszer (1991)). A large number of approximation algorithms exist, but the computational cost is still often substantial. In addition, algorithms for filtering the resulting model may be complex, and the resulting total computational demand may be very large. In addition, the best predictive model with no restrictions on size may contain only rules which are very specific, and filtering will therefore not uncover rules of a more general nature, as they are not present in the model from the outset.

ROUGH DATA MODELING

Rough Data Modeling is a method introduced by Kowalczyk (1998) which attempts to address two common problems found in traditional data mining methods: the computation cost of model generation, and the inability to tailor the method to the specific needs of each data mining session. Kowalczyk argues, as do many others, that knowledge discovery should be looked upon as an iterative process. With the large number of alternatives found at each stage in the knowledge discovery process (feature selection, discretization, data mining, etc.), it is impossible to find general guidelines which will always produce the best model. One would like to generate as large a number of models as possible, and search among these for a satisfying model. However, the computational cost of many commonly used algorithms prohibit the generation of numerous models, and in addition, most algorithms are specifically designed to maximize (or minimize) a certain measure, or a predefined combination of measures (accuracy, specificity, misclassification cost, etc.).

Rough data modeling, simplifies the model generation process, and makes it feasible to search among a large number of models. In addition, it allows the user to tailor the data mining process to suit his own needs, by allowing the user to specify in detail how to evaluate a model. As in traditional rough set theory, the starting point is a decision system $\mathcal{A} = (U, A, \{d\})$. By selecting a subset $B \subseteq A$ of the attributes in $\mathcal{A}$, the indiscernibility relation $\text{IND}(B)$ induces a set of *equivalence classes* E^B, where all elements in each class E_i^B have the same values for the attributes in B. Now, by defining a *class decision function* $\bar{d}_B : E^B \to V_d$, all objects in each equivalence class are assigned the same decision value, producing a deterministic decision system $\mathcal{A}'$. Now, a rough data model is a triple

$$\mathcal{M} = < B, \bar{d}_B, \preceq > \tag{1}$$

where B is an attribute subset of A which induces a set of equivalence classes E^B, $\bar{d}_B$ a class decision function for the equivalence classes E^B, and $\preceq$ a linear ordering of the equivalence classes in E^B. Both the class decision function and the ordering on the classes are decided by the user. According to Kowalczyk and Slisser (1997), the process of generating high-quality models using rough data modeling is a three-step process:

1. Specify the performance measure that is to be optimized, and specify the measure used to rank classes in a model.
2. Determine the search space, i. e. the collection of models which should be searched for a model maximizing the measure found in step 1. This step also involves determining the class decision function $\bar{d}_B$.
3. Determine the search procedure.

Now, each equivalence class in a particular rough data model is uniquely identified by the values of the attributes in B, and is thus equivalent to a single rule of the form

$$(a_1(x) = v_1) \wedge \ldots \wedge (a_n(x) = v_n) \to (d(x) = \bar{d}_B([x]_B)) \tag{2}$$

Where $B = \{a_1, \ldots a_n\}$, and $\{v_1, \ldots v_n\}$ are the characteristic values of the equivalence class of x, $[x]_B$.

As the characteristics used to rank each class are generated using only the objects in the particular class, and it is possible to find the equivalence classes E^B by a single run through the data set, the complexity of rough data modeling is linear with respect to the number of objects in the data set. The size of the search space is decided by the user-specified upper and lower bounds on how many attributes to include in the model, and if we search among all possible models, the expected complexity is

$$T(n) = \sum_{m=min\_attribs}^{max\_attribs} \frac{k!}{m!(k-m)!} \cdot O(n) \tag{3}$$

ROUGH MODELING

Rough data modeling is closely related to several other strategies for extracting small, yet accurate models from possibly large data sets. Rough classifiers (Piasta and Lenarcik (1996)) and feature subset search (Kohavi and Frasca (1994)) are the most closely related approaches, and while these are being explored, rough set theory is also explored, and new concepts are introduced. It is desirable to develop a framework for mining compact, yet accurate models which is able to incorporate new developments in rough set theory, such replacement of the discernibility relation with a similarity or tolerance relation (Słowiński and Vanderpooten (1997); Skowron et al. (1996)), or the incorporation of ordinal properties of attribute values obtained through the use of a dominance relation (Greco et al. (1998)).

As a step towards such a framework, we introduce the concept of *rough models*, which is a generalization of rough data models. A rough model consists of an attribute subset $B \subseteq A$ and a set of object classes (not necessarily equivalence classes) which cover the universe U, imposing no restrictions on how these classes are found. The classes may be disjoint equivalence classes obtained through the indiscernibility relation with respect to B, $\mathrm{IND}(B)$. They may also be overlapping classes obtained using a similarity relation, or obtained using the dominance relation presented by Greco et al. (1998). In addition, we replace the class decision function $\overline{d_B} : E^B \to V_d$ with an *object decision function*, $\hat{d}_B : U \to V_d$, which operates on each object instead of an entire object class. This allows us the added flexibility of being able to assign different decision values for objects which belong to the same decision class, if we so desire.

For a decision system $\mathcal{A} = (U, A, \{d\})$, a rough model is a tuple

$$M = < B, \hat{d}_B, E, \preceq, \mathcal{R} > \tag{4}$$

where $B \subseteq A$ is a set of attributes, $\hat{d}_B$ is an object decision function, E is a set of object classes which cover the universe U, $\preceq$ is a linear ordering on the classes in E, and $\mathcal{R}$ is a set of reducts for the decision system $\mathcal{A}'$, which is obtained by replacing the decision values of each object $x \in U$ with $\hat{d}_B(x)$.

Rough data models are a kind of rough model where $\mathcal{R} = B$, and $E = U/\mathrm{IND}(B)$. In addition, one can easily define for example *rough similarity models*, where the classes E are induced by a form of similarity relation (Słowiński and Vanderpooten (1997))[1], or *rough dominance models*, where the dominance relation, explained in Greco et al. (1998), is used to partition the universe into classes E.

IMPLEMENTING ROUGH MODELING

The ROSETTA system (Øhrn et al. (1998)) is a system for analyzing data using rough set methods. It has been developed as a joint effort between the Institute of Mathematics at the University of Warsaw, and the Knowledge Systems Group at the Norwegian University of Science and Technology (NTNU). A public version is available from the ROSETTA home page, (Øhrn (1999)). A large number of algorithms for reduct computation, as well as discretization, completion, classification and the other subtasks in the knowledge discovery process have been implemented as part of the ROSETTA toolkit, and we have implemented the rough data modeling approach to model generation in this system. The TRANCE system, implemented by Kowalczyk (1998), is a set of MatLab procedures which form a toolbox for performing rough data modeling. Using MatLab as a foundation for model search means that there is great flexibility in the fact that functions may be added during the process, allowing a high degree of flexibility and tailorability. However, this assumes the presence of a highly skilled MatLab user, and as the MatLab program is geared towards highly accurate mathematical computations, it's computations are unnecessarily complex.

The ROSETTA system, comprised of compiled C++ code, does not allow the same degree of flexibility, as only the preprogrammed quality and ranking functions may be used. However, due to its graphical user interface (GUI), it does not require any special knowledge from the user, and provided that the algorithms used are of equal quality, the compiled program will run faster than interpreted MatLab procedures.

In addition to implementing a simple exhaustive search of the model space, a genetic algorithm which searches the space of models has been implemented. Encoding each rough model as a binary string is simply done by constructing a bit string of length equal to the number of attributes in $\mathcal{A}$, where a value of 1 at position i means that attribute a_i is a part of the model. It is believed that the search space of all possible rough models is well suited to a genetic algorithm. It is non-continuous, making iterative-improvement algorithms ill-suited, and finding heuristics which indicate which attributes to include is very difficult, as indicated by Imam (1996).

Several of the proposed variants within the rough modeling framework were implemented. This includes the process of rough data modeling, as well as the generation of *rough Holte models*, which are identical to rough data models, except that while the attribute subset B in a rough data model is equivalent to a single reduct, each attribute $a_i \in B$ forms a reduct in a rough Holte model. This means that the resulting rules will be *univariate* rules (rules with an antecedent consisting of a single descriptor), hence the name.

Following the recommendations put forth by Provost et al. (1998), as well as experience gained through data analysis at the Knowledge Systems Group at NTNU, accuracy was discarded as a measure of the predictive power of a classifier. Instead, ROC curves (Hanley and McNeil (1982)) were used, and the area under the ROC curve (AUC) was used as the numerical measure indicating predictive capability.

[1] In this case, the classes will overlap.

RESULTS

In order to investigate the performance of the rough modeling approach within the ROSETTA toolkit, several data sets were analyzed using rough modeling, as well as traditional rough set methods for comparison. These experiments do not enable us to form any strong conclusions about the merits and deficiencies of rough modeling versus other rough set methods, but they serve as an indication. All results shown below were computed using the genetic algorithm for rough model search, as initial experiments proved this algorithm on average used only 20-20% of the time used by an exhaustive search, and still consistently returned the same models.

On the mushrooms data set, taken from the UCI machine learning repository (Blake et al. (1998)), the task of learning turned out to be too easy to shed any insight into the performance of rough models versus other methods. AUC values of 0.998 were easily obtained using rough data modeling, and the performance was therefore not investigated further. However, the rules found using rough modeling were compared to the rule set described by Duch et al. (1997), which is considered to be the smallest rule set describing the data perfectly. This rule set consists of four rules, and the rule sets found by using rough modeling were found to be very similar to the ideal rule set, both in terms of number of rules (about 10), and attributes used.

Rough models were also generated from the Pima Indian diabetes data set (Blake et al. (1998)), and the acute appendicitis data set (Hallan et al. (1997b)). The diabetes data, being part of the UCI machine learning repository, has been used in a large number of published studies. Kohavi and Sommerfield (1995) analyzed the data using a variety of algorithms, and reported accuracies of between 0.68 and 0.76, while Smith et al. (1988) reported sensitivity and specificity of 0.76 using the ADAP learning algorithm. No published AUC values for the data are known. The acute appendicitis data set has been analyzed using logistic regression (Hallan et al. (1997b,a)), obtaining AUC values of 0.920. Using dynamic reduct computation (Bazan et al. (1994)), Carlin et al. (1998) analyzed the data set, and reported a mean AUC value (and standard deviation) of 0.923 (0.023). The parameters used to generate this rule set are known, and it is thus possible to use this rule set as a benchmark with which to compare the rough models. Details, as well as a detailed description of the data set, including the discretization used for the numerical attributes, can be found in Carlin et al. (1998).

In order to investigate the effect of excluding the smallest equivalence classes from the rough data models, the size threshold for inclusion into the rough model was varied, and a rough model generated from 67% of the objects in the data set. The resulting model was used to classify the test set, consisting of 33% of the objects, and the AUC value was calculated using the trapezoidal method of approximating integrals. The results on the acute appendicitis and diabetes data are shown in Table 3.

In addition to a decision function $\hat{d}_B$ which sets the decision value for each object to the dominating decision value for the equivalence class of that object, a decision function which copies the original decision value for each object was implemented. This breaks with the principle put forth by Kowalczyk that each equivalence class should have a single decision value associated with it, and produces a model which may contain indeterministic rules. The performance of the models generated when keeping the original decision values is shown in Table 3. The same splits as for the dominating decision results were used, in order to facilitate comparison of the AUC values.

The performance of rough Holte models was only briefly investigated on the appendicitis data. The complete set of univariate rules for the data set, denoted `1R`, was used as a benchmark, and a 3-fold split-validation was carried out, with different lower limits on the classes to be included in the model. The results are collected with the other results in Table 3.

ANALYSIS

The results on the appendicitis data set indicate that the rough data models perform somewhat poorer than the optimal model generated using traditional rough set methods. In order to investigate this further, the Hanley-McNeil test for comparing correlated AUC values (Hanley and McNeil (1982)) was used to compare the measured AUC values. This was only done for the appendicitis data, as no known benchmark existed for the diabetes data. The resulting p-values from the Hanley-McNeil test are shown in Table 1. Interpreting the results from the Hanley-McNeil test is not straightforward. For the models generated using a dominant decision function, the optimal RS model seems to perform better at a significance level of $\alpha = 0.05$ for a size limit of 0, and $\alpha = 0.10$ for a size limit of 5 and 15. For a lower limit on class size of 10, no statements about significance can be made. For models generated by copying the original decision values, the only claims to significance can be made at size limits of 0 and 15, where the difference is significant at $\alpha = 0.05$.

Overall, the results from the appendicitis data indicate that rough data models perform somewhat worse than the best known RS model, but the difference is not significant if the correct threshold for class size is selected. This means that one may accept a slight drop in performance, but the drop is often insignificant. In order to examine the increase in descriptiveness, Table 2 lists the size of the different rough data models for the appendicitis data set. These numbers largely speak for themselves, while it certainly is possible to systematically analyze and investigate a model containing 800+ rules, it is a very time-consuming task. On the other hand, a model containing 10-15 rules, where each rule uses

		Size Limit			
	Decision func.	**0**	**5**	**10**	**15**
Split 1	dominating	0.0180	0.0972	0.0775	0.0001
	original	0.0414	0.3584	0.7668	0.0065
Split 2	dominating	0.0154	0.0899	0.1570	0.0528
	original	0.0156	0.9167	0.0375	0.0483
Split 3	dominating	0.0100	0.0906	0.0112	0.0122
	original	0.0181	0.8677	0.2498	0.0122

Table 1: Statistical analysis of the results on the appendicitis data. The AUC values for the models using the dominating and original decision function, listed in Table 3 were compared to the AUC value of the best rule set induced using rough set methods.

	No. rules in model		
Size limit	**Split 1**	**Split 2**	**Split 3**
RS model	893	872	851
0	107	70	77
5	17	18	17
10	9	10	8
15	5	5	5

Table 2: The number of rules in each RDM with varying size limit. The number of rules in the best RS rule set for the particular split is included for comparison.

3-4 attributes, is easily inspected. In addition, by accepting the slight performance drop from using a dominating decision functions, all the rules in the RDM will be deterministic (have a consequent consisting of a single indicated decision value). On all other data sets examined, the size of the models found by a rough model search were comparable to the results reported for the appendicitis set (between 5 and 20 rules, if a small cut-off on class size is used).

Regarding the rough Holte models, no significant differences were found between the different rough Holte models and the `1R` rule set, or between the different rough Holte models. This means that it is possible to mimic the performance of the `1R` rule set (which is comparable to the performance of the best reduct-based models) using only a handful of attributes.

DISCUSSION

The results obtained so far are not a solid enough foundation on which to make strong claims about the performance of rough models versus models mined using traditional rough set methods. There is some indication that the performance of rough models falls slightly short of the performance of larger rough set induced models, but the evidence is not conclusive. However, there is overwhelming evidence supporting the conjecture that rough data models of comparable performance to traditional RS models are far more descriptive. No universally agreed upon measure of descriptiveness exists, but a decrease in size from thousands of rules to between five and twenty, few attributes used, and well as the option of determinism, add up to a very large improvement in descriptive capability.

It is somewhat unfair to compare rough models to unfiltered models created from reducts (or approximations of reducts), but even after filtering, the size of the RS models is unlikely to be as low as that of the rough models. Even if it is, the time used to first approximate reducts, then employ one or more filtering strategies, is likely to be greater than the time spent searching for rough models. An overview of filtering strategies, as well as an investigation of the performance of heavily filtered rule sets, can be found in Ågotnes (1999).

The results in the experiments performed support the observations made by Holte (1993), that very simple rules often perform just as well as more complex rules. For the appendicitis data, it may seem that the predictive capabilities of univariate rules match, and possibly exceed those of rough data models, which contain rules with several attributes in the antecedent. However, univariate rules are less interesting from a knowledge discovery standpoint, as the only represent simple correlations between individual attribute values and a decision value.

The simplicity of rough model generation means that rough modeling is well-suited to large databases, and well-suited as an initial analysis approach. By searching for various forms of rough models, models which are of a high descriptive and predictive quality may be generated quickly. The insight gained from inspecting the rough models may then be used to process the data before using reduct-based model inducers to mine the best possible predictive model, if this is of interest.

		AUC (SD)			
Model / Data set	**Size limit**	**Split 1**	**Split 2**	**Split 3**	**Mean**
RDM Dominating Appendicitis	`RS` model	0.923 (0.028)	0.891 (0.031)	0.908 (0.031)	**0.907 (0.031)**
	0	0.804 (0.047)	0.768 (0.050)	0.776 (0.050)	**0.783 (0.049)**
	5	0.856 (0.040)	0.807 (0.046)	0.842 (0.042)	**0.835 (0.043)**
	10	0.850 (0.041)	0.834 (0.043)	0.765 (0.051)	**0.816 (0.045)**
	15	0.710 (0.056)	0.795 (0.048)	0.807 (0.046)	**0.771 (0.050)**
RDM Original Appendicitis	0	0.829 (0.044)	0.764 (0.051)	0.790 (0.048)	**0.794 (0.048)**
	5	0.891 (0.034)	0.887 (0.035)	0.914 (0.030)	**0.897 (0.033)**
	10	0.914 (0.030)	0.792 (0.048)	0.866 (0.038)	**0.857 (0.039)**
	15	0.813 (0.046)	0.806 (0.046)	0.799 (0.048)	**0.806 (0.047)**
RDM Dominating Diabetes	0	0.678 (0.038)	0.708 (0.035)	0.691 (0.035)	**0.692 (0.036)**
	5	0.666 (0.038)	0.728 (0.034)	0.689 (0.036)	**0.694 (0.036)**
	10	0.702 (0.037)	0.784 (0.032)	0.712 (0.034)	**0.732 (0.035)**
	15	0.690 (0.037)	0.733 (0.034)	0.721 (0.034)	**0.715 (0.035)**
RDM Original Diabetes	0	0.726 (0.036)	0.741 (0.034)	0.778 (0.034)	**0.748 (0.034)**
	5	0.754 (0.035)	0.774 (0.032)	0.664 (0.036)	**0.730 (0.034)**
	10	0.768 (0.034)	0.762 (0.033)	0.760 (0.033)	**0.763 (0.033)**
	15	0.771 (0.034)	0.779 (0.032)	0.750 (0.033)	**0.767 (0.033)**
RHM Original Appendicitis	`1R` model	0.904 (0.032)	0.913 (0.030)	0.906 (0.033)	**0.908 (0.032)**
	0	0.944 (0.024)	0.919 (0.029)	0.912 (0.031)	**0.925 (0.028)**
	5	0.915 (0.030)	0.871 (0.038)	0.896 (0.034)	**0.894 (0.034)**
	10	0.892 (0.034)	0.896 (0.033)	0.865 (0.039)	**0.884 (0.035)**
	15	0.914 (0.030)	0.896 (0.033)	0.898 (0.034)	**0.903 (0.032)**

Table 3: Results from experiments carried out on the appendicitis and diabetes data. The `RS` model is the best reduct-based model reported by Carlin et al. The `1R` rule set is the set of univariate rules for all attributes.

REFERENCES

T. Ågotnes. Filtering large propositional rule sets while retaining classifier performance. Master's thesis, Norwegian University of Science and Technology (NTNU), 1999.

J.G. Bazan, A. Skowron, and P. Synak. Dynamic reducts as a tool for extracting laws from decision tables. In Z.W. Raś and M. Zemankova, editors, *Methodologies for Intelligent Systems*, number 869 in Lecture Notes in Artificial Intelligence, pages 346–355. Springer-Verlag, 1994. Proceedings, 8th International Symposium, ISMIS'94, Charlotte, USA.

C. Blake, E. Keogh, and C.J. Merz. UCI repository of machine learning databases, 1998. http://www.ics.uci.edu/∼mlearn/MLRepository.html.

U. Carlin, J. Komorowski, and A. Øhrn. Rough set analysis of patients with suspected acute appendicitis. In B Bouchon-Meunier and R.R. Yager, editors, *Proceedings, Seventh Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'98, Paris, France*, pages 1528–1533. Edisions, E.D.K., 1998.

W. Duch et al. Extraction of crisp logical rules using constrained backpropagation networks. In *Proceedings, European Symposium on Artificial Neural Networks, ESANN'97*. D Facto Publications, 1997.

S. Greco, B. Matarazzo, and R. Słowiński. New developments in the rough set approach to multi-attribute decision analysis. *Bulletin of International Rough Set Society*, 2(2/3):57–87, 1998.

S. Hallan, A. Åsberg, and T.-H. Edna. Additional value of biochemical tests in suspected acute appendicitis. *European Journal of Surgery*, 163(7):533–538, 1997a.

S. Hallan, A. Åsberg, and T.-H. Edna. Estimating the probability of acute appendicitis using clinical criteria of a structured record sheet: The physician against the computer. *European Journal of Surgery*, 163(6):427–432, 1997b.

J.A. Hanley and B.J. McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143:29–36, 1982.

R.C. Holte. Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, 11:63–91, 1993.

A. Øhrn. The ROSETTA home page, 1999. http://www.idi.ntnu.no/∼aleks/rosetta/.

A. Øhrn, J. Komorowski, A. Skowron, and P. Synak. The design and implementation of a knowledge discovery toolkit based on rough sets – the ROSETTA system. In *Rough Sets in Knowledge Discovery 1: Methodology and Applications Polkowski and Skowron (1998)*, chapter 19, pages 376–399. Physica-Verlag, 1998.

I.F. Imam. An empirical study on the incompetence of attribute selection criteria. In Z.W. Raś and M. Michalewicz, editors, *Foundations of Intelligent Systems*, number 1079 in Lecture Notes in Artificial Intelligence, pages 458–467. Springer-Verlag, 1996. Proceedings, 9th International Symposium, ISMIS'96, Zakopane, Poland.

R. Kohavi and B. Frasca. Useful feature subsets and rough set reducts. In T.Y. Lin and A.M. Wildberger, editors, *3rd International Workshop on Rough Sets and Soft Computing (RSSC '94), San Jose, USA*, 1994.

R. Kohavi and D. Sommerfield. Feature subset selection using the wrapper method: Overfitting and dynamic search space topology. In U.M. Fayyad and R. Uthurusamy, editors, *Proceedings of the First International Conference on Knowledge Discovery and Data Mining, KDD'95*. AAAI Press, 1995.

W. Kowalczyk. Rough data modelling : a new technique for analyzing data. In *Rough Sets and Knowledge Discovery 1: Methodology and Applications Polkowski and Skowron (1998)*, chapter 20, pages 400–421. Physica-Verlag, 1998.

W. Kowalczyk and F. Slisser. Modelling customer retention with rough data models. In J. Komorowski and J. Zytkow, editors, *Principles of Data Mining and Knowledge Discovery*, number 1263 in Lecture Notes in Artificial Intelligence, pages 4–13. Springer-Verlag, 1997. Proceedings, First European Symposium, PKDD'97, Trondheim, Norway.

Z. Pawlak. *Rough Sets — Theoretical Aspects of Reasoning about Data*. System Theory, Knowledge Engineering and Problem Solving. Kluwer Academic Publishers, 1991.

Z. Piasta and A. Lenarcik. Rule induction with probabilistic rough classifiers. Technical report, Warszaw University of Technology, 1996. ICS Research Report 24/96.

L. Polkowski and A. Skowron, editors. *Rough Sets in Knowledge Discovery 1: Methodology and Applications*, volume 1 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag, 1998.

F. Provost, T. Fawcett, and R. Kohavi. *The Case Against Accuracy Estimation for Comparing Induction Algorithms*, 1998. To appear in *Machine Learning*.

A. Skowron, L. Polkowski, and J. Komorowski. Learning tolerance relations by boolean descriptors, automatic feature extraction from data tables. In S. Tsumoto, S. Kobayashi, T. Yokomori, H. Tanaka, and A. Nakamura, editors, *Proceedings of the Fourth International Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery, RSFD'96, Tokyo, Japan*, pages 11–17, 1996.

A. Skowron and C. Rauszer. The discernibility matrices and functions in information systems. In R. Słowiński, editor, *Intelligent Decision Support Systems — Handbook of Applications and advances in Rough Set Theory*, pages 331–362. Kluwer Academic Publishers, 1991.

R. Słowiński and D. Vanderpooten. Similarity relation as a basis for rough approximations. In P.P. Wang, editor, *Advances in Machine Intelligence & Soft Computing*, volume 4, pages 17–33. Duke University Press, 1997.

J.W. Smith, J.E. Everhart, W.C. Dickson, and W.C. Knowler. Using the adap learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the Symposium on Computer Applications and Medical Care*, pages 261–265. IEEE Press, 1988.