Rough Set Analysis of Patients with Suspected Acute Appendicitis

Ulf S. Carlin, Jan Komorowski[*]**and Aleksander Øhrn**[†]
Knowledge Systems Group
Dept. of Computer and Information Science
Norwegian University of Science and Technology
N-7034 Trondheim, Norway
{usc, janko, aleks}@idi.ntnu.no

Abstract

A data set describing 257 patients with suspected acute appendicitis is analyzed with rough set tools. The data set has previously been analyzed using logistic regression [5, 6], and the difference in performance between the two methods is found to be very small. The rough set approach additionally offers a set of decision rules, that can be interpreted without familiarity with sophisticated statistical concepts.

1 Introduction

Acute appendicitis is one of the most common problems in clinical surgery in the western world [3, 4], and the diagnosis is sometimes difficult even for experienced surgeons. Furthermore, two types of diagnostic errors have to be considered in the decision-making process: Unnecessary operations are clearly desirable to avoid, but a delayed diagnosis may lead to perforation of the appendix. Since perforation of the appendix leads to morbidity and occasionally death, a high rate of unnecessary surgical interventions is usually accepted. Analysis of collected data with the objective of improving various aspects of diagnosis is therefore potentially valuable.

Rough set theory [8, 9] is a fairly new knowledge discovery technique that has been previously applied to the medical domain (see for instance, [11, 12, 13, 16]). One advantage of the rough set approach is that a set of readable if-then rules is produced. Such rules have a potential to reveal new medical insight by pointing out strong patterns in the data material, and may also collectively function as a classifier for unseen cases.

This paper summarizes a rough set analysis of a data set describing patients with suspected acute appendicitis. The objective was to develop rules that could predict either the presence or absence of acute appendicitis on the basis of observed patient attributes. The data set has previously been studied using logistic regression [5, 6]. A comparison between both methods of analysis is done, and different aspects of their respective strengths and weaknesses are discussed.

The structure of the paper is as follows. In Sect. 2 basic notions of Pawlak's information systems and some more recent notions are recalled. The data material is presented in Sect. 3. Processing methodology is described in Sect. 4. Section 5 gives the results of processing and, finally, Sect. 6 offers an analysis and a discussion of the results.

2 Theoretical Preliminaries

An *information system* is a pair $\mathbf{A} = (U, A)$, where U is a non-empty finite set of *objects* called the *universe* and A is a non-empty finite set of *attributes* such that $a : U \to V_a$ for every $a \in A$. The set V_a is called the *value set* of a. A *decision table* is any information system of the form $\mathbf{A} = (U, A \cup \{d\})$, where $d \notin A$ is a special attribute called the *decision attribute*. The elements of A are called *condition attributes*.

Letting $\mathbf{A} = (U, A)$ be an information system, then for any $B \subseteq A$ is associated an equivalence relation $IND(B)$ as define below. $IND(B)$ is called the *B-indiscernibility relation*. If $(x, x') \in IND(B)$, then objects x and x' are indiscernible from each other by attributes from B.

$$IND(B) = \{(x, x') \in U^2 \mid \forall a \in B, a(x) = a(x')\}$$

A *reduct* is a minimal set of attributes $B \subseteq A$ such that $IND(B) = IND(A)$. If B is minimal and $IND(B) = IND(\{d\})$, B is said to be *d-relative reduct*. A reduct B defines a *functional dependency* $B \to A \backslash B$ (respectively $B \to d$), meaning that the values of the attributes $A \backslash B$ (respectively d) can be

---

[*] On leave at: Dept. of Automatic Control, Lund Institute of Technology, Box 118, SE-222 00 Lund, Sweden

[†] On leave at: Decision Systems Group, BWH and Harvard Medical School, Dept. of Radiology, 75 Francis Street, Boston, MA 02215, USA

determined if the values for attributes B are known.

Let $\mathbf{A}$ be an information system with n objects. The *discernibility matrix* of $\mathbf{A}$ [10] is a symmetric $n \times n$ matrix with entries c_{ij} where each entry consists of the set of attributes upon which objects x_i and x_j differ.

A *discernibility function* $f_{\mathbf{A}}$ for an information system $\mathbf{A}$ is a Boolean function of m Boolean variables $a_1^*, \ldots, a_m^*$ (corresponding to the attributes $a_1, \ldots, a_m$) defined as below, where $c_{ij}^* = \{a^* \mid a \in c_{ij}\}$. The set of all prime implicants of $f_{\mathbf{A}}$ determines the set of all reducts of $\mathbf{A}$.

$$f_{\mathbf{A}}(a_1^*, \ldots, a_m^*) = \bigwedge \left\{ \bigvee c_{ij}^* \mid 1 \leq j \leq i \leq n, c_{ij} \neq \emptyset \right\}$$

If we instead construct a Boolean function by restricting the conjunction to only run over column k in the discernibility matrix, we obtain the *x_k-relative discernibility function.* The set of all prime implicants of this function determines the set of all *x_k-relative reducts* of $\mathbf{A}$. These reducts reveal the minimum amount of information needed to discern $x_k \in U$ from all other objects.

A *descriptor* is an expression (a, v) where $a \in A$ and $v \in V_a$. A *decision rule* of a decision table $\mathbf{A}$ is any expression of the form $\tau \rightarrow (d, i)$ where $i \in V_d$ and τ is a Boolean combination of descriptors. The meaning $\tau_{\mathbf{A}}$ of a Boolean combination of descriptors τ is the set of objects in U with property τ, defined inductively in the obvious way. If we restrict the antecedent of the rule to be a conjunction and the consequent to be a disjunction, minimal decision rules can be obtained by computing the (x_k, d)-relative reducts and reading off the attribute values for the attributes in the reducts and the decision attributes. The *support* of a decision rule is the cardinality of the set $\tau_{\mathbf{A}} \cap (d, i)_{\mathbf{A}}$, i.e. the number of objects that match the pattern described by the rule.

The process of computing *dynamic reducts* [1] from an information system $\mathbf{A} = (U, A)$ can be seen as combining normal reduct computation with statistical resampling techniques. A family of subsystems $\mathbf{B} = (U' \subseteq U, A)$ of varying size are randomly sampled, and reducts computed from these. This process typically produces attribute subsets that define near-functional dependencies of $\mathbf{A}$, but actual functional dependencies in some sampled subtable $\mathbf{B}$. From dynamic reducts one can generate decision rules that are *approximate* in the sense that they are inconsistent. Approximate rules are more tolerant to noise and thus reveal default decisions in some sense.

We conclude this section with a comment that the task described in this paper falls into the research area known as knowledge discovery (KDD). The overall KDD process may be broken up into several steps and phases that are iterated in a waterfall-like cycle. From a data source containing raw data, all or portions of this is selected for further processing. The selected raw data is then typically pre-processed and transformed in some way, before being passed on to the data mining algorithm itself. The output patterns from the computational mining procedure are then post-processed, interpreted and evaluated, hopefully revealing new knowledge previously buried in the data. Along the way, backtracking on each of the steps will in practice inevitably occur. For a thorough discussion of these issues the reader can consult [14, 15]. Here we shall discuss only some details specific to the data set.

3 Data Material

The data set studied in this paper consists of 257 patients with suspected acute appendicitis. For each patient the attributes listed in Tab. 1 and Tab. 2 were recorded. The binary and numerical attributes are summarized in Tab. 1 and Tab. 2, respectively. For further details about the collection of the data material, see [5, 6].

For each patient the surgeon estimated the patient's risk of having acute appendicitis in increments of 10% from 0 to 100%. There were nine different surgeons with two to six years of surgical training who participated in this probability estimation. The estimation was done before the result of a blood test was ready. The attributes based on the blood test, and thus not available to the surgeon when he performed the probability estimation, are: ESR, CRP, WBC, and NEUTRO. The probability estimates were, of course, neither a part of the logistic regression analysis nor of the rough set analysis.

The DIAGNOSIS attribute is the (*a posteriori*) decision attribute d in the analysis. It shows which patients actually turned out to have appendicitis. As can be seen in Tab. 1, 98 patients (38%) turned out to have appendicitis and 159 (62%) turned out to have some other disease or non-specific abdominal pain. The final diagnosis of acute appendicitis was based on histological examination of the excised appendix. Other diagnoses were based on routine investigation with repeated clinical examination, biochemical tests, imaging techniques and, if necessary, surgery.

In the analysis, different subsets of all the attributes will be used as $\mathbf{A}$ in the decision system $\mathbf{A} = (U, A \cup d)$. This is done in order to make the comparison of the diagnostic ability of the logistic regression model, the rough set model, and the surgeons' probability estimate as fair as possible wrt. to the attributes.

Table 1: Binary attributes

Attribute	Description	Statistics Yes % (count)	 No % (count)
SEX	Male sex?	55.3 (142)	44.7 (115)
ANOREXIA	Anorexia?	69.3 (178)	30.7 (79)
NAUSEA	Nausea or vomiting?	70.8 (182)	29.2 (75)
PREVIOUS	Previous surgery?	9.3 (24)	90.7 (233)
MOVEMENT	Aggravation of pain by movement?	61.5 (158)	38.5 (99)
COUGHING	Aggravation of pain by coughing?	59.9 (154)	40.1 (103)
MICTUR	Normal micturition?	87.2 (224)	12.8 (33)
TENDRLQ	Tenderness in right lower quadrant?	86.0 (221)	14.0 (36)
REBTEND	Rebound tenderness in right lower quadrant?	55.3 (142)	44.7 (115)
GUARD	Guarding or rigidity?	30.7 (79)	69.3 (178)
CLASSIC	Classic migration of pain?	49.4 (127)	50.6 (130)
DIAGNOSIS	(Final diagnosis:) acute appendicitis?	38.1 (98)	61.9 (159)

Table 2: Numerical attributes

Attribute	Description	Unit	Statistics Mean (SD)	 Median	 Range
AGE	Age	years	26.8 (17.0)	22	3–86
DURATION	Duration of pain	hours	35.3 (53.8)	22	2–600
TEMP	Rectal temperature	°C	37.8 (0.746)	37.7	36.4–40.3
ESR	Erythrocyte sedimentation rate	mm	14.1 (15.8)	10	1–90
CRP	C–reactive protein concentration	mg/L	32.8 (48.7)	12	0–260
WBC	White blood cell count	$\times 10^9$	12.3 (4.79)	12.1	2.9–31.0
NEUTRO	Neutrophil count	%	77.1 (11.4)	80	38–93

The rough set theory is based on the concept of indiscernibility. Numerical attributes should be discretized into intervals, so that numbers falling within the same interval are deemed as being indiscernible. The numerical attributes in Tab. 2 are discretized in Tab. 3.

Several algorithms for automatic discretization exist, but none was used in this analysis. Instead, the discretization was done manually. The CRP, WBC, and NEUTRO attributes were discretized by a medical expert, while the ESR attribute was discretized using the same intervals as in [7]. The AGE, DURATION, and TEMP attributes were discretized into three intervals, each containing approximately the same number of objects.

4 Methodology

A typical setting for KDD experiments is shown in Fig. 1. A group of objects to be studied is split in two groups. A discovery algorithm is applied on the first group and a classifier is returned. The classifier is then tested on the other group of objects and some kind of performance measure is computed. The process is typically iterated by, for instance, varying the splitting in a systematic fashion, and/or adjusting various parameters in the discovery process.

If the rough set approach is used in this general scheme, the box denoted “Training” in Fig. 1 can be decomposed as shown in Fig. 2. Discretization *may* be done before the splitting of the main data table in Fig. 1, as is done in this paper. The usual approach when standard discretizations of the attribute domains are not available, is to discretize the training set (e.g. with an automatic algorithm), and then use the resulting discretization on the test set. The first step in Fig. 2 is to compute reducts from the discretized table. This can be done with any type of reducts. In the present analysis full discernibility dynamic reducts are computed. After reduct computation, rules are generated from the reducts and the decision table. Each rule may have an associated certainty measure indicating the strength of the rule. After the steps of reduct computation and rule generation, “weak” reducts and rules can be filtered. This is done with the current data

Table 3: Discretization of numerical attributes

Attribute	Intervals	Count	Description
AGE	$[-\infty, 17)$	84	Low
	$[17, 31)$	90	Middle
	$[31, \infty)$	83	High
DURATION	$[-\infty, 13)$	96	Short
	$[13, 30)$	71	Middle
	$[30, \infty)$	90	Long
TEMP	$[-\infty, 37.5)$	84	Low
	$[37.5, 38.1)$	89	Middle
	$[38.1, \infty)$	81	High
ESR	$[-\infty, 10)$	126	Normal
	$[10, 25)$	99	Slightly raised
	$[25, \infty)$	32	Considerably raised
CRP	$[-\infty, 6)$	103	Normal
	$[6, 40)$	95	Slightly raised
	$[40, \infty)$	59	Considerably raised
WBC	$[-\infty, 10.0)$	91	Normal
	$[10.0, 14.0)$	80	Slightly raised
	$[14.0, \infty)$	86	Considerably raised
NEUTRO	$[-\infty, 75)$	87	Normal
	$[75, 85)$	87	Slightly raised
	$[85, \infty)$	83	Considerably raised

set in [2]. In the rough set approach, the classifier in Fig. 1 is a set of rules.

In the case of the study by Hallan et al. [5, 6], the "Training" box in Fig. 1 contained a logistic regression analysis, which resulted in a function that maps each patient to a probability of disease. Hallan et al. chose a splitting strategy where the original group of objects were split in two approximately equal parts. Logistic regression analysis was then performed on one part and subsequently tested on the other. This was done for 20 random splits, and from each iteration an ROC curve[1] was generated by varying the cut-off probability of disease. The mean area under the 20 ROC curves was subsequently calculated.

We performed rough set analysis in a similar fashion, i.e. with 20 different splits of the original table. These splits were chosen randomly, and are thus not the same splits as those used by Hallan et al. ROC curves were generated as follows. When a set of rules is applied to an object, several rules may fire. Furthermore, these rules may indicate different decision classes and may have different support counts associated with them. (If no rules fire, a default class may be chosen.) One way to resolve such conflicts is through voting. An election procedure is simulated among the firing rules, where each rule gets a number of votes according to its support. Typically the decision class with the largest number of votes would be selected, but it is possible to prioritize a decision class by letting all objects with a voting percentage for class C above some threshold be classified as such. By varying this prioritization threshold, points on an ROC curve can be generated. The area under the ROC curve (AUC) can then be computed using the trapezoidal integration rule, and the mean AUC can be computed over a number of different runs where different splits of the original table are used.

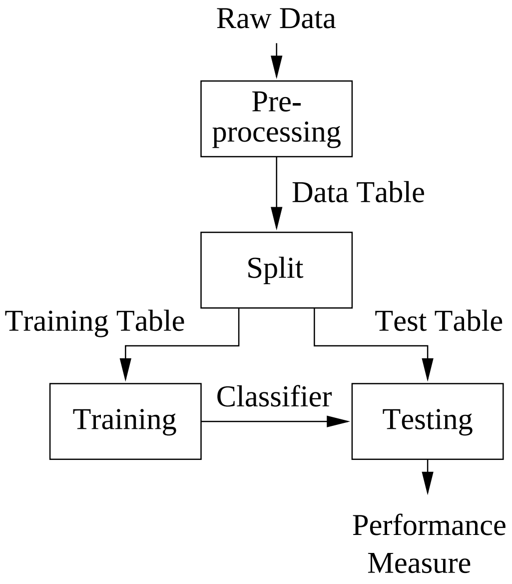

Figure 1: General training/testing cycle

Hallan et al. presented in [6] analyses on three differ-

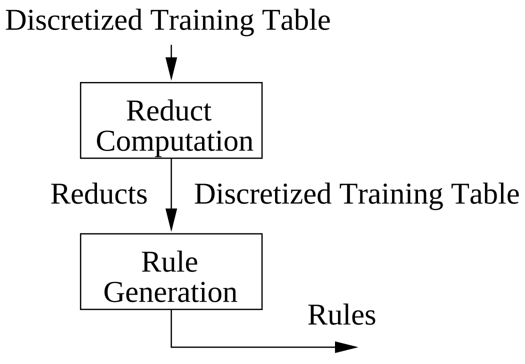

Figure 2: Training in the rough set approach

---

[1] An ROC curve is a plot of *1 - specificity* on the x-axis versus *sensitivity* on the y-axis for different cut-off values. Sensitivity (resp. specificity) is the proportion of the patients with positive (resp. negative) disease status who are correctly identified by the test.

Table 4: Results: Average AUC ± SD

Variable set	Logistic Regression	Rough Sets	Surgeons
A	0.854 ± 0.0283	0.850 ± 0.0235	
B	0.901 ± 0.0174	0.905 ± 0.0231	
C	0.920 ± 0.0238	0.923 ± 0.0225	
Clinical variables			0.817

ent variable sets. The first variable set, which will be called A for simplicity, consists of the following clinical variables: CLASSIC, REBTEND, SEX, TENDRLQ, COUGHING, GUARD. The second analysis used the variable WBC in addition, let $B = A \cup \{WBC\}$. The last analysis used variables CRP and NEUTRO in addition. This variable set will be called C, $C = B \cup \{CRP, NEUTRO\}$. Hallan et al. found that adding other clinical variables or the ESR did not improve the logistic regression model further.

All rough set computations were carried out using the Rosetta software system [15]. The presented rough set analyses were done on the same sets of variables as used by Hallan et al. for reasons of comparison. Full discernibility dynamic reducts were calculated using an exhaustive algorithm on each sampled subtable. The dynamic reduct sampling strategy was the following: Subtables were sampled on 10 equally spaced levels with 50 samples per level from 5% to 95% of the original table. Then rules were generated from all of the resulting reducts.

5 Results

The results of the rough set analysis and the logistic regression analysis by Hallan et al. are summarized in Tab. 4. The results are presented as average AUC values ± standard deviation for experiments on 20 different splits of the original table as described in Sect. 4. We see that the logistic regression models and the rough set models perform approximately equally well. Both methods perform somewhat better than the surgeons. Just like the logistic regression model on variable set C, the corresponding rough set model did not improve further when adding clinical variables or the ESR.

In Tab. 5, a comparison of the rough set approach and the surgeons is done using some other common measures of performance (sensitivity, specificity and accuracy), in addition to the mean area under the ROC curves. The two following methods are added. The strategy of the best attribute, CLASS, is: If classic migration of pain is present at a patient, classify as "sick", else classify as "healthy". The baseline strategy is to classify all patients as the majority decision class, which in this data table is "not appendicitis". Any advanced method should at least perform better than those two strategies.

Table 5: Comparison between rough set approach and surgeons.

Method	Sens.	Spec.	Acc.	AUC
RS on A	0.684	0.837	0.774	0.850
RS on B	0.801	0.853	0.830	0.905
RS on C	0.876	0.850	0.858	0.923
Surgeons	0.867	0.679	0.751	0.817
Best attribute	0.755	0.667	0.700	0.711
Baseline	0	1	0.619	0.5

We can see from this table that the rough set approach scores slightly better than the surgeons on accuracy, when only clinical variables are considered. The accuracy rises from 0.774 to 0.858 when CRP, WBC, and NEUTRO are added to the analysis. The high sensitivity and low specificity for the surgeons reflects the usual tendency to overestimate the diagnosis acute appendicitis to avoid perforation. Using the prioritization threshold mentioned in Sect. 4, the rough set models could also have been tuned to overestimate the diagnosis, giving a higher sensitivity and a lower specificity. Such tuning also affects the accuracy, but not the AUC.

6 Analysis and Discussion

Rough set analysis confirms the results obtained in [6] and it adds the possibility of inspecting the discovered diagnostic algorithm. On the other hand, logistic regression is a well-known and extensively used statistical method which will be preferred by most medical doctors. The relative advantage of rough sets needs to be further investigated in order to show substantial gains over the standard approaches, if it is to be widely used in the medical community. On the other hand, computer scientists trained in the rough set methods are likely to provide services of higher quality and possibly lead to the discovery of new medical knowledge.

The results obtained by rough sets are nevertheless remarkable. Using only 257 objects the diagnostic algorithm provides an explicit representation and performs better than a surgeon with a 2 to 6 year training. Thus such algorithms may eventually be very useful decision aids, even for the experienced diagnosticians.

In order to get a better assessment of the relative "goodness" of both techniques, their predictive capability should be assessed on a new independent set of data. There is a risk that both are over-optimistic.

7 Acknowledgments

Special thanks to Stein Hallan, MD, for making the data available. This research is supported in part by the European Union 4th Framework Telematics project CardiAssist, by the Human Capital and Mobility Norwegian Research Council contract #101341/410, by the Norwegian Research Council grant #74467/410, and by the Norwegian Research Council grant for Cooperation with Central Europe.

References

[1] J. Bazan, A. Skowron, P. Synak (1994), Dynamic Reducts as a Tool for Extracting Laws from Decision Tables, Proc. Symp. on Methodologies for Intelligent Systems, Charlotte, NC, USA, October 16–19, Lecture Notes in Artificial Intelligence, Springer Verlag, Vol. 869, pp. 346–355.

[2] U. Carlin (1998), Mining medical data with rough sets, M.Sc. (siv.ing.) thesis, Department of Computer and Information Science, The Norwegian University of Science and Technology, Trondheim, Norway.

[3] S. Eriksson (1996), Acute Appendicitis – Ways to Improve Diagnostic Accuracy, European Journal of Surgery, Vol. 162, pp. 435–442.

[4] J. M. Grönroos, J. J. Forsström, K. Irjala, T. J. Nevalainen (1994), Phospholipase A_2, C-Reactive Protein, and White Blood Cell Count in the Diagnosis of Acute Appendicitis, Clinical Chemistry, Vol. 40, No. 9, pp. 1757–1760.

[5] S. Hallan, A. Åsberg, T.-H. Edna (1997), Estimating the Probability of Acute Appendicitis Using Clinical Criteria of a Structured Record Sheet: The Physician Against the Computer, European Journal of Surgery, Vol. 163, No 6, pp. 427–432.

[6] S. Hallan, A. Åsberg, T.-H. Edna (1997), Additional Value of Biochemical Tests in Suspected Acute Appendicitis, European Journal of Surgery, Vol. 163, No 7, pp. 533–538.

[7] W. P. Oosterhuis, A. H. Zwinderman, M. Teeuwen, et al. (1993), C reactive protein in the diagnosis of acute appendicitis, European Journal of Surgery, Vol. 159, pp. 115–119.

[8] Z. Pawlak (1982), Rough Sets, International Journal of Computer and Information Sciences, Vol. 11, No 5, pp. 341–356.

[9] Z. Pawlak (1991), Rough Sets – Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht.

[10] A. Skowron and C. Rauszer (1992), The Discernibility Matrices and Functions in Information Systems, In R. Slowinski (ed.), Intelligent Decision Support – Handbook of Applications and Advances of the Rough Sets Theory, Dordrecht, Kluwer Academic Publishers, pp. 331–362.

[11] K. Slowinski (1988), Rough Set Approach to Analysis of Data from Peritoneal Lavage in Acute Pancreatitis, Medical Informatics, 13, no. 3, pp. 143–159.

[12] K. Slowinski (1992), Rough classification of HSV patients, In R. Slowinski (ed.), Intelligent Decision Support – Handbook of Applications and Advances of the Rough Sets Theory, Dordrecht, Kluwer Academic Publishers, pp. 77–94.

[13] S. Vinterbo, L. Ohno-Machado, H. Fraser (1998), Prediction of Acute Myocardial Infarction using Rough Sets, In preparation.

[14] A. Øhrn, J. Komorowski (1997), Rosetta – A Rough Set Toolkit for Analysis of Data, Proc. Third International Joint Conference on Information Sciences, Durham, NC, USA, March 1–5, Vol. 3, pp. 403–407.

[15] A. Øhrn, J. Komorowski, A. Skowron, P. Synak (1998), The Design and Implementation of a Knowledge Discovery Toolkit Based on Rough Sets – The Rosetta System, To appear in Rough Sets in Knowledge Discovery, L. Polkowski and A. Skowron (Eds.), Physica Verlag, 24 pages.

[16] A. Øhrn, S. Vinterbo, P. Szymanski, J. Komorowski (1997), Modelling Cardiac Patient Set Residuals using Rough Sets, Proc. AMIA Annual Fall Symposium (formerly SCAMC), Nashville, TN, USA, October 25–29, pp. 203–207.