Published in European Journal of Operational Research, vol. 160, no. 3, 2005, 696-709

Supporting Triage of Children with Abdominal Pain in the Emergency Room

Sz. Wilk[1], R. Słowiński[1], W. Michałowski[2], S. Greco[3]

[1] Institute of Computing Science, Poznań University of Technology, Poznań, Poland
[2] Faculty of Administration, University of Ottawa, Ottawa, Canada
[3] Faculty of Economics, University of Catania, Catania, Italy

Abstract

The paper describes a methodology used for selecting the most relevant clinical features and for generating decision rules based on selected attributes from a medical data set with missing values. These rules will help emergency room (ER) medical personnel in triage (initial assessment) of children with an abdominal pain. Presented approach is based on rough set theory extended with the ability of handling missing values and with the fuzzy measures allowing estimation of a value of information brought by particular attributes. The proposed methodology was applied for analyzing the data set containing records of patients with abdominal pain, collected in the emergency room of the cooperating hospital. Generated rules will be embedded into a computer decision support system that will be used in the emergency room. The system based on results of presented approach should allow supporting of triage accuracy by the emergency room staff, and reducing management costs.

Keywords

rough sets, medicine, decision support systems, triage

1. Introduction

Abdominal pain in childhood is a highly prevalent symptom in modern society caused by a range of organic diseases, psychosocial disturbances and emotional disorders. In many cases, the exact cause is never determined. Many of these young patients when in severe pain end up in hospital emergency departments where medical staff must focus on identifying the very small portion of cases with a serious organic disease that are in need of urgent treatment. For the majority of such patients where the diagnosis of *appendicitis* is in doubt, investigations and repeated assessments conducted by different physicians are time consuming and may be painful.

There is empirical evidence that highlights an obvious advantage of the rapid triage of patients with abdominal pain (Fioravanti *et al.* 1988). However, the central difficulty of such triage is accurate initial assessment based on a limited number of clinical signs, symptoms and tests (attributes) that in combination contribute the most to the diagnosis and management. Hence, a relevantly reduced set of attributes should assist the triage nurse and help the ER physician.

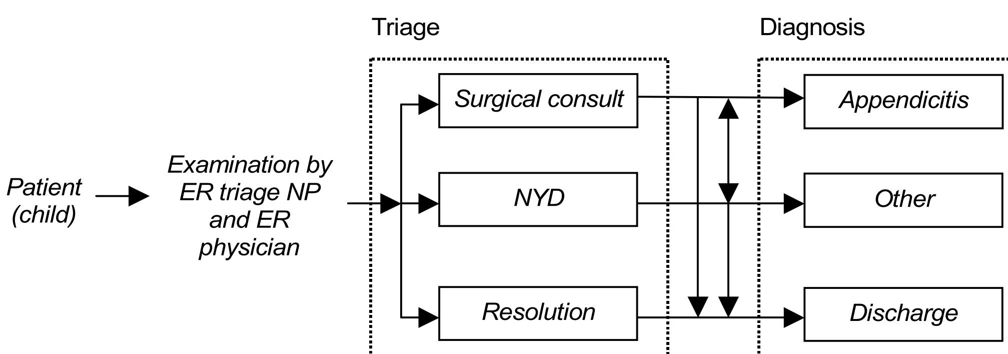
The focus of this study is on finding the set of relevant clinical symptoms and signs, and to generate decision rules based on them. These rules are designed to support emergency room (ER) staff in assessing children with abdominal pain. Created decision algorithm (i.e. a set of decision rules) will be embedded into a decision support system (DSS) that will be used in emergency rooms of hospitals participating in the study.

The paper is organized as follows. We start by describing the ER management of child with abdominal pain, and with the data set used in this study. In section 2 we present the methodology , and in section 3 we describe the experiments and their results. Finally, the last section presents conclusions.

1.1. ER management of child with abdominal pain

The typical process of ER management of a child complaining of abdominal pain is illustrated in Figure 1. It also establishes a framework for use of the decision support system, based on the results of the presented approach, in triage of patients.

Figure 1. ER management of child with abdominal pain



At present, a child is examined first by the ER triage NP (nurse practitioner) to evaluate his/her condition. This assessment is followed by a detailed examination conducted by the ER physician. The possible outcomes of this evaluation are: *resolution*, *surgical consults*, and *not-yet-diagnosed* (*NYD*). *Resolution* indicates that the abdominal pain spontaneously subsides in the absence of a clinical diagnosis. Such a patient can be discharged to the care of his/her regular physician. *Surgical consult* implies that *acute appendicitis* is suspected and a surgeon is called for further evaluation. *NYD* indicates the need for further in-hospital clinical evaluation.

Triage is the first stage in the process of patient management. The final (actual) diagnosis is known when patient is discharged, and with patients triaged as *surgical consult*, the final diagnosis is obtained from the post- surgery pathology report.

There is no consensus in the medical literature regarding the most effective management of abdominal pain patients. Several studies advocate use of some form of a scoring system as a clinical support tool (Anatol and Holder 1995; Hallan and Edna 1997), while others argue that only an experienced member of the pediatric surgical team is capable of the reliable diagnosis of *acute appendicitis* (Simpson and Smith 1996). Current practice in teaching hospitals shows an ER triage accuracy of 50-60% achieved by an ER triage NP in case of diagnosing *appendicitis,* thus there is a need to support this stage of child management by reducing a number of incorrectly triaged patients. This observation prompted us to focus our attention on the triage stage of the patient management process.

The results of our study will be used to develop a decision support system. The system will be aimed at being used by the ER medical personnel (triage NP and ER physician). There is no need to support surgeons, because usually they are able to diagnose appendicitis with very high accuracy, however their availability in the ER is limited and they cannot examine every patient.

1.2. Data set

We began our research with the development of the data set. It contains the records of 647 patients with abdominal pain seen during the 1997 – 2000 period in the ER of the Children's Hospital of Eastern Ontario (Ottawa, Ontario). For each patient the 12 attributes (clinical signs, symptoms and tests)[1] and their values as given in Table 1, were extracted from the charts. It should be noticed that not every caregiver was able to conduct all the

[1] These are the attributes recommended in current medical texts for the diagnosis of patients with abdominal pain.

examinations necessary to obtain the values for every attribute. The ER triage NP is trained to acquire the values of some of the attributes (for example elements of a medical history), while the ER physician is trained to carry out a complete examination.

Most of the attributes are nominal (e.g. gender or type of pain), some are numerical (e.g. temperature or WBCC), and some indicate a location of a condition on patient's abdomen (e.g. location of pain or location of tenderness). Values of the latter ones were collected with a help of special abdomen pictograms, on which attending ER physician marked proper location. Numerical values were discretized according to medical practice, and "location" attributes have values converted into nominal ones using an algorithm developed by surgeons and ER physicians for that specific purpose.

Table 1. Clinical attributes collected in the data set

Attribute	Description	Domain
Age	Patient's age	0–5 years, ≥ 5 years
Sex	Gender	male, female
PainDur	Duration of pain	≤ 24 hours, 1–7 days, > 7 days
PainSite	Location of pain	RLQ, lower abdomen, other
PainType	Type of pain	continuous, other
Vomiting	Vomiting occurred	yes, no
PrevVis	Previous visit to ER in last 48 hours	yes, no
Tempr	Temperature	< 37 °C, 37–39 °C, ≥ 39 °C
TendSite	Location of tenderness	RLQ, lower abdomen, other
Guarding	Localized abdominal muscle guarding	absent, present
RebTend	Localized abdominal rebound tenderness	absent, present
WBCC	White blood cell count	≤ 4000, 4000–12000, ≥ 12000

According to medical practice, patient records stored in the data set were classified into 3 classes: *resolution*, *surgical consult*, and *NYD*. The detailed partition is given in Table 2. The data set is unbalanced – 61% of objects belong to the *resolution* class (and so is the default accuracy), the *surgical consult* class includes 30% of all records, and the remaining 9% makes the *NYD*.

Table 2. Partition of patient records into decision classes

Class	# of examples	% of examples
Resolution	394	60.9
Surgical consult	195	30.1
NYD	58	9.0
Total	647	100.0

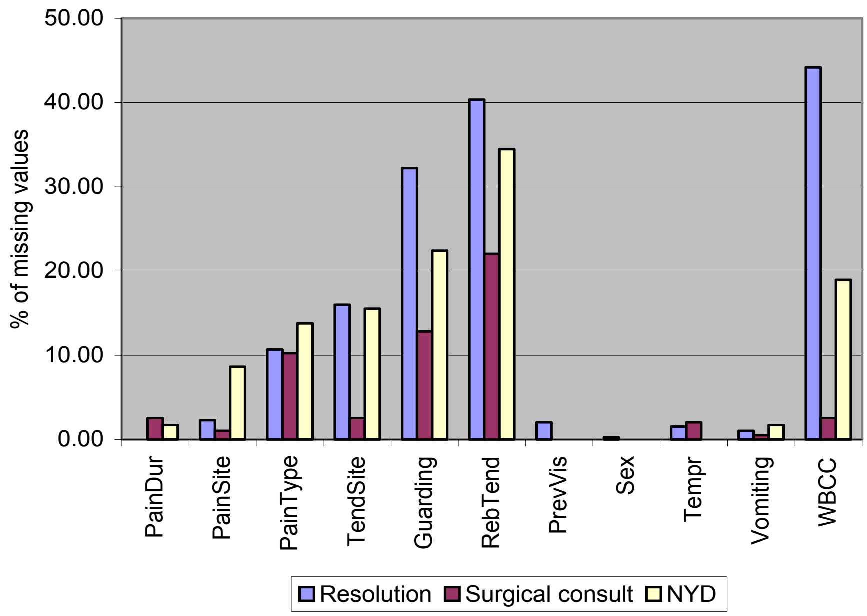
The data was collected as part of a prospective chart study and thus are not complete. Detailed information about missing values is presented in Table 3. Values of only one attribute – Age – are known for all patients. Five attributes have more than 10% of values is missing, and three of them (Guarding, WBCC, and RebTend) have more than 20% of missing values. The unusually large portion of missing values for the last attribute is caused by the fact that it was recorded only for the records collected in the year 2000.

Table 3. Missing values of attributes [in %]

		Class		
Attribute	Overall	Resolution	Surgical consult	NYD
Age	0.0	0.0	0.0	0.0
PainDur	0.9	0.0	2.6	1.7
PainSite	2.5	2.3	1.0	8.6
PainType	10.8	10.7	10.3	13.8
TendSite	11.9	16.0	2.6	15.5
Guarding	25.5	32.2	12.8	22.4
RebTend	34.3	40.4	22.1	34.5
PrevVis	1.2	2.0	0.0	0.0
Sex	0.2	0.3	0.0	0.0
Tempr	1.6	1.5	2.1	0.0
Vomiting	0.9	1.0	0.5	1.7
WBCC	29.4	44.2	2.6	19.0

Figure 2 reveals an interesting pattern of missing values within the decision classes. Except of two attributes (PainDur, and Tempr), the largest ratio of missing data appears for the *resolution* class, and the smallest for the *surgical consult*. This may be explained by the fact, that if, after performing a few basic examinations the most likely triage is *resolution*, then other symptoms and signs are not checked. Clearly patients triaged as *surgical consult* or *NYD* are examined more thoroughly.

Figure 2. Distribution of the missing values in decision classes



An important issue here is how to treat the attributes with significant number of missing values. They should not be discarded, as their importance clearly depends on a decisional context. The WBCC attribute is a very good example of such a situation. It has 44% of missing values for the *resolution* class, but less than 3% for the *surgical consult* class. One can conclude here that triage of almost half of the patients was so obvious that it did not require any further investigations and that ordering the WBCC would be a waste of time and the resources. Thus, the WBCC in this particular context is not an important attribute. On the other hand, when a patient is triaged as the *surgical consult*, the results of the WBCC are very important for the correct diagnosis[2]. It seems that the ratio of missing values provides a "time stamp" of the examination – larger amount of missing data is associated with the symptoms and signs considered later in a management process.

[2] This 3% of missing values should be most likely attributed to a fact that the WBCC was done but its results were not properly recorded on a patient's chart in the ER.

Published in European Journal of Operational Research, vol. 160, no. 3, 2005, 696-709

2. Methodology

The data set created from the patients' charts was analyzed for regularities using rough set theory and fuzzy measures. In this section we describe the basic notions behind rough set theory and its extensions (Pawlak 1982, 1991; Pawlak and Słowiński 1994; Słowiński 1995), and also the fuzzy measures (Grabish 1997; Greco *et al.* 1998; Shapley 1953) that was used for additional evaluation of the attributes.

2.1. Rough set theory

For rough set analysis the data is supplied in the form of an *information table*, in which rows represent objects (patients' charts) and columns represent attributes (clinical signs, symptoms and tests, and triage outcomes recorded on the charts). Each cell of the table indicates an evaluation (quantitative or qualitative) of an object represented by the corresponding row by means of an attribute represented by the corresponding column. Formally, the information table is the 4-tuple $S = \langle U, Q, V, f \rangle$, where U is a finite set of objects (universe), Q is a finite set of attributes, $V = \bigcup_{q \in Q} V_q$ and V_q is a domain of an attribute q, f: $U \times Q \rightarrow V$ is a function called *information function* such that $f(x,q) \in V_q$ for each $q \in Q$, $x \in U$. The set Q is divided into a set C of *condition* attributes, and a set D of *decision* attributes (for the data set described in this paper it is a singleton – the triage outcome).

Each object x of U is described by a vector of evaluations (attribute values), called *description* of x in terms of the attributes from Q. This vector represents available information about x. Objects having identical description are called *indiscernible*. In general, the *indiscernibility relation* on U, denoted by I_p, is associated with every (non-empty) subset of attributes $P \subseteq Q$

$$I_P = \{(x, y) \in U \times U : f_q(x) = f_q(y), \forall q \in P\}. \qquad (1)$$

Clearly, the relation (1) is an equivalence relation (reflexive, symmetric and transitive); thus, it partitions set U into equivalence classes called *P-elementary sets.* The family of all the equivalence classes of relation I_P is denoted by $U \mid I_P$ and the equivalence class containing an object $x \in U$ is denoted by $I_p(x)$. If $(x,y) \in I_P$, then objects x and y are *P-indiscernible*.

The partitions of set U induced by subsets of condition attributes and subsets of decision attributes represent *knowledge* about U.

Published in European Journal of Operational Research, vol. 160, no. 3, 2005, 696-709

The key idea of rough sets is related to approximation of knowledge expressed by decision attributes using knowledge that is expressed by condition attributes. Rough set theory answers several questions related to such approximation:

(a) Is the data contained in the information table consistent?

(b) What are the non-redundant subsets of condition attributes (*reducts*) ensuring the same quality of classification as the whole set of condition attributes?

(c) What are the condition attributes (*core*) that cannot be eliminated from the approximation without decreasing the quality of the approximation?

(d) What minimal "*if ..., then ...*" decision rules can be induced from the approximations?

Several important aspects of rough set theory made it of particular interest to the researchers evaluating real data sets describing various decision situations (Lin and Cercone 1997). With respect to the input information, rough sets allow us to analyze both quantitative and qualitative data, and inconsistencies need not be removed prior to the analysis, as they are dealt with by separating certain and doubtful knowledge extracted from the information table. With reference to the output information, rough sets allow us to acquire *a posteriori* information regarding the relevance of a particular attribute (or subsets of attributes) for the quality of classification. Moreover, the final result presented in the form of "*if..., then...*" decision rules that are based on the most relevant attributes, is easy to interpret.

For a demonstration of the basic concepts of rough sets, let us assume that X is a non-empty subset of U, for example an equivalence class with respect to set D of decision attributes, and $P \subseteq C$ is a subset of condition attributes.

We say that object $x \in X$ *belongs certainly* to X if all objects from the P-elementary set $I_p(x)$ also belong to X, i.e. $I_p(x) \subseteq X$. Then, for a given P, information about object x is *consistent* with information about other objects from U.

We say, moreover, that object $x \in U$ *could belong* to X if at least one object from the P-elementary set $I_p(x)$ belongs to X, i.e. $I_p(x) \cap X \neq \emptyset$. If $\emptyset \neq I_p(x) \cap X \neq I_p(x)$ then, for a given P, information about object x is *inconsistent* with information about other objects from $I_p(x)$.

For $P \subseteq C$, the set of all objects belonging certainly to X constitutes the *P-lower approximation* of X, denoted by $\underline{P}(X)$, and the set of all objects that could belong to X constitutes the *P-upper approximation* of X, denoted by $\overline{P}(X)$:

$$\underline{P}(X) = \{x \in U : I_P(x) \subseteq X\}, \tag{2}$$

$$\overline{P}(X) = \{x \in U : I_P(x) \cap X \neq \emptyset\}. \tag{3}$$

The difference between the upper and lower approximations of X is called the P-boundary of X:

$$Bn_P(x) = \overline{P}(X) - \underline{P}(X). \tag{4}$$

The P-boundary of X is composed of inconsistent objects that belong to X with some ambiguity. The following relations hold: $\underline{P}(X) \subseteq X \subseteq \overline{P}(X)$, $\underline{P}(X) = U - \overline{P}(U - X)$.

The family of all the sets $X \subseteq U$ having the same lower and upper approximations is called a *rough set*.

Rough approximations of a subset $X \subseteq U$ can be extended to partitions of U; in particular to the partition induced by decision attributes from D. This partition corresponds to the classification of objects into decision classes –lower approximations of decision classes represent certain knowledge, upper approximations represent possible knowledge, and the boundaries represent doubtful knowledge about the classification, expressed in terms of condition attributes from $P \subseteq C$.

Given a partition of U into decision classes, $\boldsymbol{Cl} = \{Cl_t, t \in T\}$, $T = \{1,...,n\}$, the P-boundary with respect to $k>1$ classes $\{Cl_{t1},\ldots,Cl_{tk}\} \subseteq \{Cl_1,\ldots,Cl_n\}$ is defined as

$$Bd_P(\{Cl_{t1},...,Cl_{tk}\}) = \left(\bigcap_{t=t1,...,tk} Bn_P(Cl_t)\right) \cap \left(\bigcap_{t \neq t1,...,tk} (U - Bn_P(Cl_t))\right). \tag{5}$$

The objects from $Bd_P(\{Cl_{t1},...,Cl_{tk}\})$ can be assigned to one of the classes $Cl_{t1},\ldots,Cl_{tk}$, however, P and all its subsets do not provide enough information to do this assignment precisely.

Using the rough approximations of decision classes, it is possible to induce *decision rules* describing the classification represented by examples contained in the information table. These are logical statements (implications) of the type "*if ..., then...*" where the antecedent (condition part) is a conjunction of the elementary conditions concerning particular condition attributes and the consequence (decision part) is a disjunction of possible assignments to particular classes of a partition of U induced by decision attributes. Given a partition $\boldsymbol{Cl}$ of U, the syntax of the rule is the following:

Published in European Journal of Operational Research, vol. 160, no. 3, 2005, 696-709

$$\text{"}\textit{if } f(x,q_1) = r_{q1} \ \textit{ and } f(x,q_2) = r_{q2} \ \textit{ and } \ldots f(x,q_p) = r_{qp}, \\ \textit{then } x \text{ is assigned to } Cl_{t1} \ \textit{ or } \ldots Cl_{tk}\text{"}, \tag{6}$$

where $\{q_1,...,q_p\} \subseteq C$, $(r_{q1},...,r_{qp}) \in V_{q1}\times...\times V_{qp}$ and $\{Cl_{t1},...,Cl_{tk}\} \subseteq \{Cl_1,...,Cl_n\}$. If the consequence is univocal, i.e. k=1, then the rule is *exact*, otherwise it is *approximate* or *uncertain*.

Let us observe that for any $Cl_t \in \{Cl_1,...,Cl_n\}$ and $P \subseteq C$, the definition (2) of P-lower approximation of Cl_t can be rewritten as:

$$\underline{P}(Cl_t) = \{x\in U\text{: for each } y\in U, \text{ if } yI_Px, \text{ then } y\in Cl_t\}. \tag{7}$$

Thus, the objects belonging to the lower approximation $\underline{P}(Cl_t)$ can be considered as prototypes for induction of exact decision rules.

Therefore, the statement "*if* $f(x,q_1) = r_{q1}$ *and* $f(x,q_2) = r_{q2}$ *and* ... $f(x,q_p) = r_{qp}$, *then* x is assigned to Cl_t", is accepted as an exact decision rule if and only if there exists at least one object $y\in\underline{P}(Cl_t)$, $P=\{q_1,\ldots,q_p\}$, such that $f(y,q_1) = r_{q1}$ and $f(y,q_2) = r_{q2}$ and $\ldots f(y,q_p) = r_{qp}$.

Given $\{Cl_{t1},\ldots,Cl_{tk}\} \subseteq \{Cl_1,...,Cl_n\}$ we can write

$$Bd_P(\{Cl_{t1},\ldots,Cl_{tk}\}) = \{x\in U\text{: for each } y\in U, \textit{ if } yI_Px, \\ \textit{then } y\in Cl_{t1} \textit{ or } \ldots Cl_{tk}\}. \tag{8}$$

Thus, the objects belonging to the boundary $Bd_P(\{Cl_{t1},\ldots,Cl_{tk}\})$ can be considered as a basis for induction of approximate decision rules.

The analysis of large information tables shows that the calculation of approximations according to (2) and (3) may result in large P-boundary of X. Consequently, it leads to weak decision rules (supported by few objects from lower approximations). In such a case it seems reasonable to relax the conditions for the assignment of objects into lower approximations by allowing it to include some inconsistent objects. This relaxation is called a *variable precision model* (VPM) and is described in (Ziarko 1993). The VPM defines lower approximations using a limited number of counterexamples that is controlled by a pre-defined level of certainty β ($0 < \beta \leq 1$). In the VPM, the P-lower approximation of X in U is defined as:

$$\underline{P}(X) = \left\{ x \in U : \frac{|I_P(x) \cap X|}{|I_P(X)|} \geq \beta \right\}. \tag{9}$$

If β is set to 1, then the VPM operates in the same way as in definition (2).

Decision rules induced from the lower approximations of decision classes defined by (7) have univocal consequences (decisions); however, *confidence index* of each rule (defined as the number of objects matching both the condition and decision part of the rule to the number of objects matching the condition part only) varies from β to 1.

The *quality of the approximation* of $X \subseteq U$ by the attributes from P is given as the ratio:

$$\gamma_P(X) = \frac{|\underline{P}(X)|}{|X|}, \tag{10}$$

where $0 \leq \gamma_P(X) \leq 1$ and the quality represents the relative frequency of the objects correctly classified by means of the attributes from P.

The *quality of the approximation of classification* $\boldsymbol{Cl}$ by set of attributes P is given as:

$$\gamma_P(\boldsymbol{Cl}) = \frac{\sum_{i=1}^{n} |\underline{P}(Cl_i)|}{|U|}. \tag{11}$$

It is called in short the *quality of classification* and specifies the ratio of all P-correctly classified objects to all objects in the information table.

Each minimal subset $P \subseteq C$ such that $\gamma_P(\boldsymbol{Cl}) = \gamma_C(\boldsymbol{Cl})$ is called a *reduct* of S and denoted by $RED_{\boldsymbol{Cl}}(C)$. The information table can have more than one reduct. The intersection of all reducts is called a *core* and is denoted by $CORE_{\boldsymbol{Cl}}(C)$. The core is composed of the indispensable attributes that cannot be removed from the information table without decreasing the quality of classification. Condition attributes that do not belong to any reduct are called superfluous.

2.2. Missing values

The classical rough set theory is not well suited for dealing with missing values. If such values appear in the data set, an additional preprocessing is required to convert them. Commonly used preprocessing methods are (Grzymała-Busse and Ming 2000):

- Coding of missing values with a special value (i.e. N/A) and then treating them as known ones,

- Replacing missing values by particular known ones (for example by average or most frequent values in the whole data set or in the considered decision class),
- Replacing each incomplete object by several artificially created, with missing values replaced by all possible combinations of known ones.

All mentioned approaches have serious shortcomings. When using the first one, it is possible to obtain decision rules having conditions based on N/A values. Such rules may be difficult to interpret and use. The other methods may falsify the original data, especially when the number of missing values is large (then it may be even impossible to create all combinations of known values to replace missing ones).

There are several approaches that extend rough set theory to handle data containing missing values directly, without any preprocessing. Some of them are based on the similarity relation (Kryszkiewicz 1998a, 1998b), others on the valued tolerance relation (Stefanowski, Tsoukias 2001). The extension outlined below uses modified indiscernibility relation (Greco *et al.* 2001); it was chosen for robustness of decision rules induced from rough approximations, which is particularly important in the context of medical applications.

The definition of the information table ($S = \langle U, Q, V, f\rangle$) is extended by assuming that the set V is augmented to include the missing value (indicated by "*").

Instead of the indiscernibility relation I_P, a new type of relation, denoted by I_P^* is introduced. For each object $x, y \in U$ and for each subset of attributes $P \subseteq Q$, $y\, I_P^*\, x$ means that $f(x, q) = f(y, q)$, or $f(x, q) = *$, or $f(y, q) = *$, for every $q \in P$. Let $I_P^*(x) = \{y \in U\colon y\ I_P^*\ x\}$ for each $x \in U$ and for each $P \subseteq Q$. I_P^* is reflexive and symmetric but not transitive binary relation. Finally, let $U_P^* = \{x \in U\colon f(x, q) \neq *$ for at least one $q \in P\}$.

Using I_P^* the definitions of the P-lower and P-upper approximation of X become:

$$\underline{I}_P^*(X) = \left\{x \in U_P^* : I_P^*(x) \subseteq X\right\}, \tag{12}$$

$$\overline{I}_P^*(X) = \left\{x \in U_P^* : I_P^*(x) \cap X \neq \varnothing\right\}. \tag{13}$$

The P-lower approximation can be also calculated using the VPM as

$$\underline{I}_P^*(X) = \left\{x \in U_P^* : \frac{\left| I_P^*(x) \cap X \right|}{\left| I_P^*(x) \right|} \geq \beta\right\}. \tag{14}$$

Published in European Journal of Operational Research, vol. 160, no. 3, 2005, 696-709

The approximations defined in (12) and (13) are further used to calculate the *P*-boundary of *X*, accuracy of approximation of *X*, and the quality of the approximation of *X*.

Given the partition ***Cl*** of *U*, one can calculate the quality of the classification of ***Cl*** and use this measure to find the *reducts* and the *core* of attributes.

Using the rough approximations (12) and (13), it is possible to induce a generalized description of the examples contained in the information table in terms of *decision rules* (see 2.1).

Since each decision rule (6) is an implication (similar to (7) for I_P^*), a *minimal* decision rule represents a unique implication in the sense that there is no other implication having a subset of elementary conditions and the same consequent.

We say that $y \in U$ *supports* the exact decision rule "*if* $f(x,q_1) = r_{q1}$ *and* $f(x,q_2) = r_{q2}$ *and* ... $f(x,q_p) = r_{qp}$, *then* x is assigned to Cl_j", if [$f(y,q_1) = r_{q1}$ and/or $f(y,q_1) = *$] and [$f(y,q_2) = r_{q2}$ and/or $f(y,q_2) = *$] ... and [$f(y,q_p) = r_{qp}$ and/or $f(y,q_p) = *$] and $y \in Cl_t$.

Similarly, we say that $y \in U$ *supports* the approximate decision rule "*if* $f(x,q_1) = r_{q1}$ *and* $f(x,q_2) = r_{q2}$ *and* ... $f(x,q_p) = r_{qp}$, *then* x is assigned to Cl_{t1} *or* ... Cl_{tk}", if [$f(y,q_1) = r_{q1}$ and/or $f(y,q_1) = *$] and [$f(y,q_2) = r_{q2}$ and/or $f(y,q_2) = *$] ... and [$f(y,q_p) = r_{qp}$ and/or $f(y,q_p) = *$] and $y \in Bd_C^*(\{Cl_{t1},\ldots,Cl_{tk}\})$.

Decision rules induced from lower approximations defined by (14) have univocal consequences (decisions); however, the confidence index of each rule (see 2.1) varies from β to 1.

2.3. The fuzzy measures

The quality of classification (calculated using either I_P or I_P^*) satisfies the properties of the set functions called *fuzzy measures*. Such measures can be used for modeling the importance of coalitions (Grabisch 1997), or as proposed in (Greco *et al*. 1998) to assess the relative value of the information supplied by each attribute and to analyze interactions among the attributes (using the quality of classification calculated according to (11)). Let us explain this point in greater detail.

Let $C=\{q_1,...,q_n\}$ be a finite set, whose elements could be the players in a game, condition attributes in an information table, different criteria in a multicriteria decision problem, etc. Let $PS(C)$ denote the power set of C, i.e. the set of all subsets of C. A *fuzzy measure* on C is a set function μ: $PS(C)\rightarrow[0,1]$ satisfying the following axioms:

Published in European Journal of Operational Research, vol. 160, no. 3, 2005, 696-709

a) $\mu(\emptyset) = 0$, $\mu(C) \leq 1$,

b) $A \subseteq B$ implies $\mu(A) \leq \mu(B)$, for all $A, B \in PS(C)$.

Within the game theory, the function $\mu(A)$ is called the characteristic function and represents the payoff obtained by the coalition $A \subseteq C$ in a cooperative game (Shapley 1953); in the multi-attribute classification $\mu(A)$ can be interpreted as the conjoint importance of attributes from $A \subseteq C$.

In the game theory some indices were proposed as specific solutions of cooperative games. One of the most important is the *Shapley value* (Shapley 1953), defined for every element $q_i \in C$ as:

$$\phi_S(q_i) = \sum_{K \subseteq C - \{q_i\}} \frac{(n - |K| - 1)!|K|!}{n!} [\mu(K \cup \{q_i\}) - \mu(K)]. \qquad (15)$$

The Shapley value can be interpreted as an average contribution of the element q_i to all the possible coalitions (combinations) of elements from C.

For ***Cl*** being a partition of U, $\mu(K) = \gamma_K(\boldsymbol{Cl})$ for every $K \subseteq C$; the value of $\mu(C)$ is shared among elements of C, i.e.

$$\sum_{i=1}^{n} \phi_S(q_i) = \gamma_C(\boldsymbol{Cl}). \qquad (16)$$

Thus, the Shapley value $\phi_S(q_i)$ can be used to assess the contribution of a single attribute q_i to the quality of classification. Attributes with higher value of $\phi_S(q_i)$ are considered to better explain relationships in a data set.

3. Results

The medical data set described in section 1.2 was analyzed using the ROSE software (Prędki and Wilk, 1999). the data set contains large number of missing values, thus an approach presented in section 2.2 was used. The analysis was aimed at:

- Finding the clinical symptoms and signs that are the most relevant for classifying a patient as *resolution*, *surgical consult*, or *NYD*,
- Inducing the set of decision rules based on the attributes selected in the previous step that ensure high classification accuracy of patients in the ER. These rules, after consulting them with the surgeon, will be further used as a basis of decision support system.

Published in European Journal of Operational Research, vol. 160, no. 3, 2005, 696-709

Initial analysis identified one reduct containing all condition attributes, forcing us to use Shapley value to identify the most important attributes.

Table 4 gives the Shapley values for all condition attributes. The table also gives percentages of missing values for each attribute, and the quality of classification that was achieved by using the set containing the attributes starting from the first one (according to the sort order). For example, the quality of classification equal to 0.094 for PainSite means that it was achieved for the set containing Tempr, PainDur, Sex, Vomiting, and PainSite. The Shapley values and the quality of classification are also given in Figure 3.

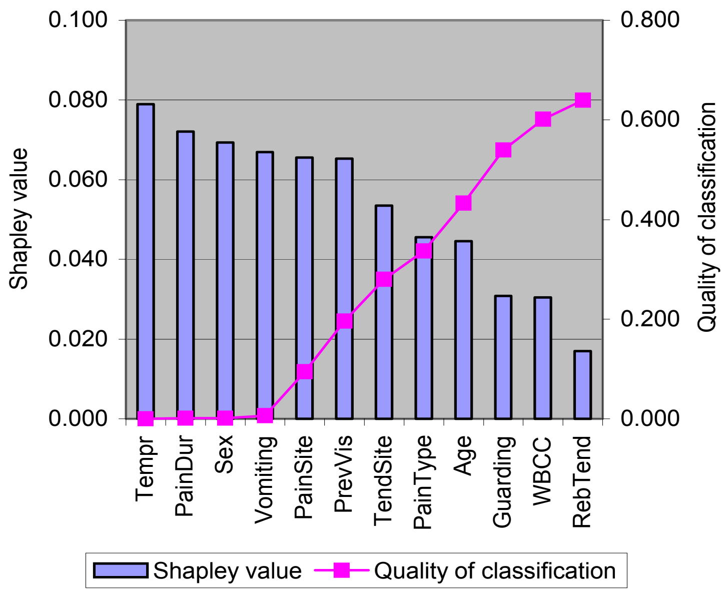
Table 4. Shapley values sorted in a descending order

Attribute	Shapley value	Quality of classification	% of missing values
Tempr	0.079	0.000	1.5
PainDur	0.072	0.002	0.9
Sex	0.069	0.002	0.2
Vomiting	0.067	0.006	0.9
PainSite	0.066	0.094	2.5
PrevVis	0.065	0.196	1.2
TendSiteSite	0.053	0.280	11.9
PainType	0.046	0.337	10.8
Age	0.045	0.433	0.0
Guarding	0.031	0.539	25.5
WBCC	0.030	0.601	29.4
RebTend	0.017	0.640	32.3

It is worth to notice that the order of the attributes in Table 4 has a close fit to the order of the percentage of missing values, especially for those that have more than 10% of missing data. It is consistent with the intuition, that less information the attribute bears, the less relevant it is according to the Shapley index ranking. The only exception being Age, which values are given for all patients, but its low importance may be explained by very crude discretization (see Table 1) that may entail loss of information.

Published in European Journal of Operational Research, vol. 160, no. 3, 2005, 696-709

Figure 3. Shapley values and the quality of classification



For illustrative purposes we also calculated the Shapley values for the data set where missing values were replaced by unique N/A value. The results presented in Table 5 are very different from those given in Table 4. For example the WBCC attribute that is undefined for almost 30% of patients appears to be the most explanatory one. Clearly, a very questionable conclusion. We believe that such simple comparison shows the advantages of the approach based on the I_P^* relation that was used to develop Table 4. Moreover there is another problem that might arise if missing values were coded as N/A. As the WBCC attribute has missing values mainly for records from *resolution* and *NYD* classes, it is very likely that a rule pointing at one of these classes and containing the condition WBCC = N/A would be created. As this condition cannot be considered as 'WBCC is unimportant' (N/A is treated as unique value, equal only to another N/A), such rule is unacceptable because it forces not to perform the WBCC test. In reality, rules of this kind are hardly useful and it is extremely difficult to interpret them.

Table 5. Shapley values (missing values coded as NA)

Attribute	Shapley value	Quality of classification	% of missing values
WBCC	0.107	0.005	29.4
Guarding	0.097	0.005	25.5
PainType	0.085	0.014	10.8
PainSite	0.084	0.178	2.5
TendSiteSite	0.083	0.377	11.9
RebTend	0.079	0.493	34.3
Tempr	0.077	0.612	1.5
PainDur	0.074	0.725	0.9
Sex	0.074	0.835	0.2
Vomiting	0.067	0.883	0.9
PrevVis	0.055	0.903	1.2
Age	0.033	0.915	0.0

The choice of the most relevant attributes in Table 4 was not obvious, so we decided to use a set of the thresholds starting with the set containing top five attributes according to their Shapley values (Tempr, PainDur, Sex, Vomiting, and PainSite). For those attributes a considerable increase of the quality of classification was observed. Then we iteratively extended this set with the remaining attributes in the sequence resulting from descending Shapley values. RebTend was appended as the last, so we finished with the set containing all 12 condition attributes.

For each set of attributes we tested the classification accuracy of corresponding decision rules. The classification accuracy was estimated using 10-fold cross-validation tests (Mitchell, 1997). In order to get more reliable results, the validation tests were repeated five times and their results were averaged over all repetitions. The decision rules were induced using two algorithms: LEM2 (Grzymała-Busse, 1992), and Explore (Stefanowski, Vanderpooten, 2001). We selected these algorithms to compare two different perspectives of decision table description and their predictive abilities.

The LEM2 algorithm generates the minimum set of decision rules (the set does not contain any redundant rules). We used the VPM variant of the algorithm that generates rules covering all examples from lower approximations calculated using the variable precision model. In all tests we fixed the β parameter at 0.8 (see definition (9) and (14)). This value was elaborated in the experiments showing that decreased β led to stronger decision rules with satisfactory discriminating abilities.

The Explore algorithm generates the set of satisfactory rules, i.e. the rules that satisfy requirements specified by the user. The user specifies the requirements in terms of the maximum number of conditions in a rule (rule length), the minimum rule strength and the minimum confidence index (the confidence index can be seen as a counterpart of β in the VPM variant of LEM2). In our experiments we did not restrict the length of generated rules, and the minimum confidence index was set to 0.8 (to preserve consistency with LEM2). The only one parameter that we modified during the tests was the minimum rule strength. The modification procedure is described in details in (Stefanowski, Wilk, 2001). Basically we started with maximum rule strength observed for the rules generated by LEM2. Then we modify them in order to increase the classification accuracy, estimated with a single repetition of 10-fold cross validation test.

The approach that we used to handle missing values, required us to introduce some modifications into rule induction algorithms. Firstly, we assumed that a condition would be satisfied by a value equal to the value specified in the condition or by missing one (i.e. condition Vomiting = 'yes' would be met by 'yes' and by missing value). Secondly, the algorithms were altered to ensure the robustness of generated rules. A decision rule is robust if it covers at least one object that has non-missing (known) values for the attributes used in the conditions.

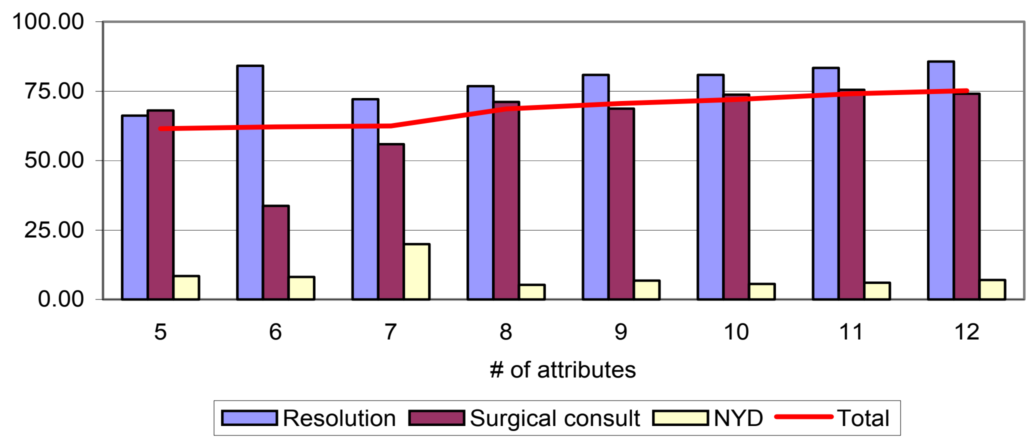
Results of cross-validation tests are presented in Table 6 and 7, and in Figure 4 and 5. The total accuracies (calculated for all decision classes) are also collected in Figure 6. The best accuracy was obtained for the rules generated by the Explore algorithm, using LEM2 led to similar accuracy, although slightly lower. However for the set of features containing all condition attributes both tested algorithms gave the same accuracy.

The rules induced by the Explore algorithm gave the best accuracy for the *surgical consult* class. Moreover, the accuracy was very stable (Figure 5). This may suggest, that it was possible to generate "accurate" rules (i.e. ensuring high classification accuracy) using the smallest subsets of condition attributes. LEM2 did not offer this kind of stability, as there was considerable decrease of accuracy for the subsets containing 6 and 7 attributes (Figure 4).

Table 6. Accuracy of classification – LEM2

	Accuracy of classification [%]			
# of attributes	Total	Resolution	Surgical consult	NYD
5	61.51	66.16	68.07	8.40
6	62.10	84.15	33.73	8.13
7	62.48	72.11	55.87	19.93
8	68.65	76.79	71.16	5.20
9	70.52	80.81	68.75	6.80
10	72.00	80.91	73.77	5.53
11	74.06	83.36	75.54	6.00
12	75.11	85.70	74.07	7.00

Figure 4. Accuracy of classification – LEM2



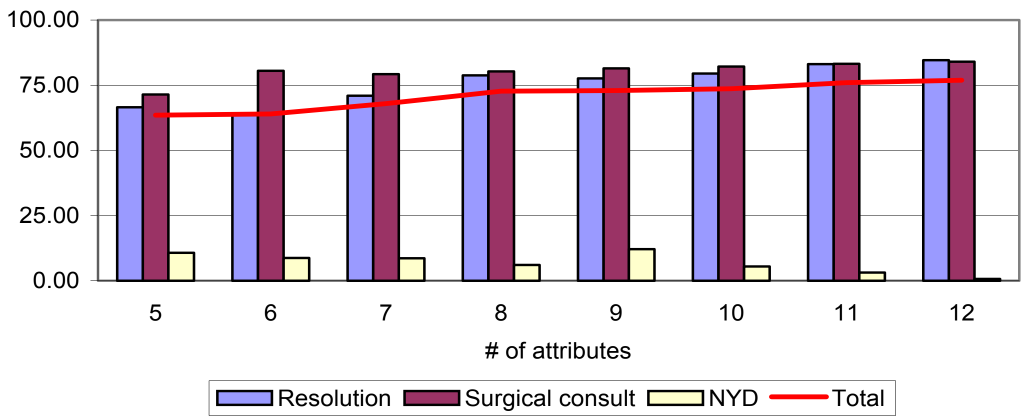
The accuracy for the *resolution* class increased as the set of analyzed attributes was extended (the only one exception was noticed for LEM2 and 6 attributes). This may suggest that additional attributes were necessary to improve the classification of the healthy patients, because, as it was mentioned above, the accuracy for the surgically consulted patients was constant.

Published in European Journal of Operational Research, vol. 160, no. 3, 2005, 696-709

Table 7. Accuracy of classification – Explore

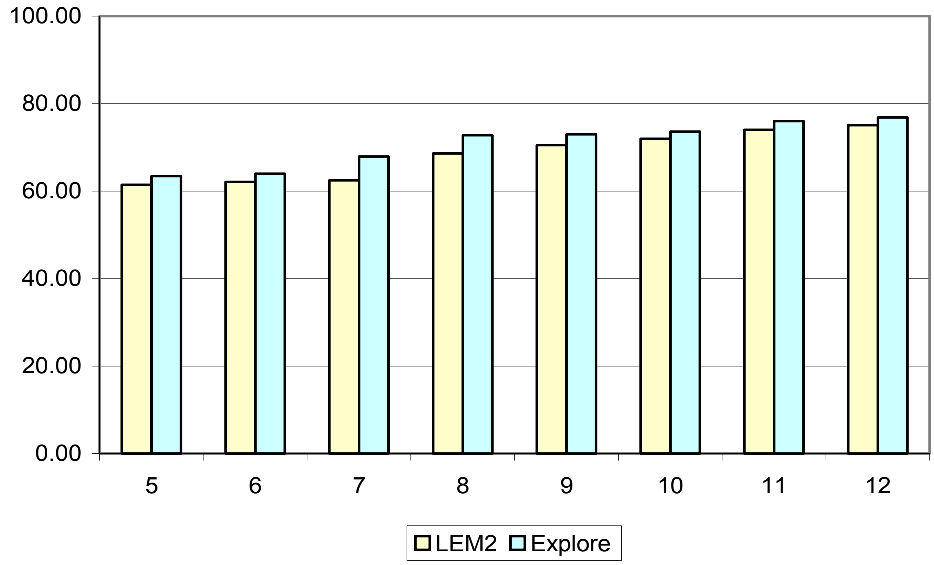
	Accuracy of classification [%]			
# of attributes	Total	Resolution	Surgical consult	NYD
5	63.46	66.57	71.48	10.67
6	64.01	64.05	80.56	8.73
7	67.91	71.00	79.30	8.60
8	72.76	78.75	80.35	6.07
9	72.95	77.67	81.47	12.07
10	73.64	79.45	82.13	5.46
11	76.01	83.13	83.25	3.13
12	76.88	84.57	83.99	0.67

Figure 5. Accuracy of classification – Explore



None of the algorithms was able to induce rules classifying the *NYD* patients with acceptable accuracy. The detailed analysis of cross-validation test results revealed that in most cases the *NYD* patients were classified to the *resolution* category. This type of misclassification is very unfavorable, as it may endanger patient's health (when a sick patient is sent back home from the ER). Such mistakes, made mainly between these two classes suggest that either not all necessary features that are normally considered by the ER staff to distinguish between healthy and *NYD* patients, were stored in the data base, or it is impossible to discern patients from the two classes with acceptable accuracy, and the ER staff decides to keep all doubtful patients in the hospital for observation. These specific misclassifications require further explanation by the surgeon.

Figure 6. Comparison of classification accuracies



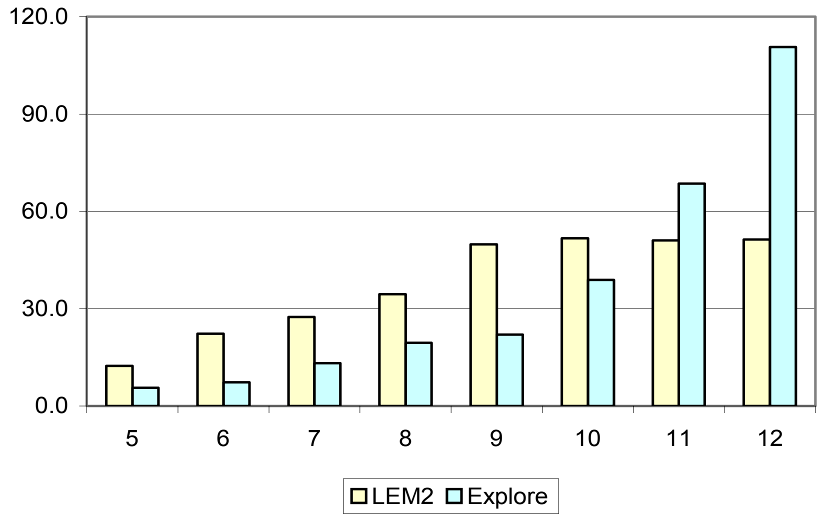
The average number of rules generated in validation tests by the tested rule induction algorithms is presented in Table 8 and in Figure 7. In most cases (except of the last two subsets of attributes) LEM2 generated more rules. It is also interesting to notice that for LEM2 the number of induced rules for the last four sets (containing 9 and more attributes) did not change. Explore produced the smaller number of rules for all, but the last two, sets of features; for the set containing all condition attributes Explore generated twice as many rules as LEM2.

Table 8. Average number of rules generated in 10-fold cross validation tests

	Average # of rules	
# of attributes	LEM2	Explore
5	12.4	5.6
6	22.3	7.3
7	27.4	13.2
8	34.4	19.5
9	49.8	22.0
10	51.7	38.8
11	51.0	68.5
12	51.3	110.6

Published in European Journal of Operational Research, vol. 160, no. 3, 2005, 696-709

Figure 7. Average number of rules generated in 10-fold cross validation tests



Finally, after consulting the surgeon, we decided to select three sets of attributes – containing 5, 8 and 11 attributes respectively. The first subset is the smallest set that gives acceptable classification accuracy and values of all included features can be collected by the ER triage NP. The other sets were selected as milestones because they are the smallest sets that contain two attributes important from the medical point of view – PainType and WBCC.

For the selected sets we induced decision rules using the Explore algorithm and the same values of parameters as in the corresponding validation tests. Obtained sets of rules, after verifying them by the surgeon and embedding into a computer DSS, will be used at various stages of handling a child in the ER, depending on the amount of available information, i.e. the rules created for 5 attributes could be used to suggest the rapid triage shortly after arriving to the ER, while the rules for 11 attributes could be applied when necessary examinations have been conducted.

Here we would like to stress, that the Explore algorithm was used for generating rules from the previous, complete (i.e., with no missing values) and much smaller version of the considered data set (175 patients' charts from two decision classes: *resolution* and *surgical consult*), and the results were promising (Michałowski *et al.*, 2001; Rubin *et al*. 2000). We generated reducts, induced decision rules from the reduced data sets and tested their classification abilities. The best set of rules achieved an accuracy of 66% in cross-validation tests, and it was positively evaluated by the surgeon.

Published in European Journal of Operational Research, vol. 160, no. 3, 2005, 696-709

Tables 9 and 10 present the rules generated for 5 and 8 attributes respectively. Because of the number of rules we decided not to include the rules created for 11 attributes. The relative strength of a rule presented in the tables is a ratio of objects covered by the rule and belonging to the class pointed by the rule to all examples from the decision class pointed by the rule (e.g. for the first rule in Table 10 the relative strength equal to 54.6%, it means that this rule covers 54.6% examples from the *resolution* class).

Table 9. Decision rules generated by Explore for 5 attributes

Diagnosis	PainDur	Site	Sex	Tempr	Vomiting	Relative strength [%]
Resolution		other				54.6
Resolution				<37	absent	38.6
Surgical consult	1 – 7 days	RLQ	male		present	28.7
Surgical consult		RLQ	male	<37	present	16.9
NYD	≤24 hours			>39		1.7
NYD		other	male	>39	absent	3.4

Published in European Journal of Operational Research, vol. 160, no. 3, 2005, 696-709

Table 10. Decision rules generated by Explore for 8 attributes

Diagnosis	PainDur	PainSite	PainType	TendSite	PrevVis	Sex	Tempr	Vomiting	Relative strength [%]
Resolution		other							54.6
Resolution				other					53.6
Resolution			intermittent				<37		38.8
Resolution			intermittent					absent	37.8
Resolution							<37	absent	38.6
Surgical consult	1 – 7 days	RLQ				male		present	28.7
Surgical consult		RLQ	constant			male		present	33.3
Surgical consult		RLQ		RLQ		male		present	34.4
Surgical consult		RLQ			no	male		present	30.3
Surgical consult		RLQ				male	<37	present	16.9
Surgical consult	1 – 7 days	RLQ	constant		no	male			33.8
Surgical consult		RLQ	constant	RLQ	no	male			41.0
Surgical consult		RLQ	constant	RLQ	no		≥37 ≤39		41.0
Surgical consult		RLQ	constant		no	male	≥37 ≤39		23.1
Surgical consult			constant	RLQ	no	male	≥37 ≤39		24.6
Surgical consult			constant	RLQ	no	male		present	31.3
Surgical consult				RLQ	no	male	≥37 ≤39	present	17.4
NYD		other		other		male	>39		6.9
NYD		other				male	>39	absent	3.4
NYD				other		male	>39	absent	3.4
NYD	24 hours		constant	other			≥37 ≤39	present	3.4

4. Conclusions

We conducted the analysis of the medical data set with missing values that was aimed at finding the set of relevant attributes and generating decision rules to classify patients in the ER.

The reduct calculated for the data set contained all condition attributes, hence we used the Shapley value as a measure of attribute's information value. The reduced sets

Published in European Journal of Operational Research, vol. 160, no. 3, 2005, 696-709

were generated in an incremental manner and for each subset we performed 10-fold cross-validation tests using the rules generated by the LEM2 and Explore algorithms. The highest classification accuracy was obtained for the latter one.

After consultations with the domain expert, we selected three subsets of attributes containing 5, 8 and 11 attributes, and generated decision rules using the Explore algorithm.

The analysis described in the paper would not be possible without employing a new approach to the evaluation of missing values. This approach allowed us to consider differential information content of the attributes, depending on a number of the missing values and their distribution among decision classes. It also allowed us to propose a modified way of matching rules, depending on the values of the attributes available at a specific stage of a triage process.

Acknowledgement

The work on this paper was supported by KBN research grant no. 8T11F 006 19, by the Foundation for Polish Science, subsidy no. 11/2001, and by the grant from the Natural Sciences and Engineering Research Council of Canada. We would like to thank Dr. Steven Rubin from the Children's Hospital of Eastern Ontario for his support and remarks on medical issues.

References

1. Anatol, T.I., Holder, Y., 1995. A scoring system for use in the diagnosis of acute abdominal pain in childhood, The West Indian Medical Journal 44 67-69.
2. Banzhaf J.F., 1965. Weighted voting doesn't work: A mathematical analysis, Rutgers Law Review 19 317-343.
3. Fioravanti M.., Di Cesare F., Ramelli L., La Torre F., Nicastro A., Messinetti S., Lazarro R., 1988. Presurgery information and psychological adjustment to enterostomy, Italian Journal of Surgical Science 18 55-61.
4. Grabisch, M., 1997. k-Order Additive Discrete fuzzy measures and their representation, Fuzzy Sets and Systems 92 167-189.
5. Greco S., Matarazzo B., Słowiński R., 1998. Fuzzy measures as a technique for rough set analysis. In: Proc. 6th European Congress on Intelligent Techniques & Soft Computing (EUFIT'98), Aachen, pp. 99-103.

6. Greco S., Matarazzo B., Słowiński R., 2001. Dealing with missing data in rough set analysis of multi-attribute and multi-criteria decision problems. In: Zanakis, S.H., Doukidis, G., Zopounidis, C. (Eds.), Decision Making: Recent Developments and Worldwide Applications, Kluwer Academic Publishers, Dordrecht/Boston, pp. 295-316.
7. Grzymała-Busse, J.W., 1992. LERS – A system for learning from examples based on rough sets. In: Słowiński, R. (Ed.), Intelligent Decision Support - Handbook of Applications and Advances of the Rough Set Theory, Kluwer Academic Publishers, Dordrecht/Boston, pp. 3-18.
8. Grzymała-Busse, J.W., Ming, H., 2001. A comparison of several approaches to missing attribute values in data mining. In: Ziarko, W., Yao, Y. (Eds.), Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, vol. 2005, Springer-Verlag, Berlin, pp. 340-347.
9. Hallan, S., Edna, T., 1997. Estimating the probability of acute appendicitis using clinical criteria of a structured record sheet: the physician against the computer, European Journal of Surgery 163 427-432.
10. Kryszkiewicz, M., 1998a. Rough set approach to incomplete information system, Information Sciences 112 39-49.
11. Kryszkiewicz, M., 1998b: Properties of incomplete information systems in the framework of rough sets. In: Polkowski, L., Skowron, A. (Eds.), Rough Sets in Data Mining and Knowledge Discovery, Physica-Verlag, Heidelberg, pp.422-450.
12. Lin, T., Cercone, N. (Eds.), 1997. Rough Sets and Data Mining, Kluwer Academic Publishers, Dordrecht/Boston.
13. Michalowski, W., Rubin, S., Słowiński, R., Wilk, Sz., 2001. Triage of the child with abdominal pain: a clinical algorithm for emergency patient management, Journal of Paediatrics and Child Health 6 (1) 23-28.
14. Mitchell, T., 1997. Machine Learning, McGraw Hill.
15. Pawlak, Z., (1982). Rough sets, International Journal of Information & Computer Sciences 11 341-356.
16. Pawlak, Z., (1991). Rough Sets. Theoretical Aspects of Reasoning about Data, Kluwer, Dordrecht.

17. Pawlak, Z., Słowiński, R., 1994. Rough set approach to multi-attribute decision analysis, European Journal of Operational Research 72 443-459.

18. Prędki, B., Wilk, Sz., 1999. Rough set based data exploration using ROSE system. In: Raś, Z.W., Skowron, A. (Eds.), Foundations of Intelligent Systems. Lecture Notes in Artificial Intelligence, vol. 1609, Springer-Verlag, Berlin, pp. 172-180 (ROSE homepage: http://www-idss.cs.put.poznan.pl/rose).

19. Rubin, S., Michalowski, W., Słowiński, R., Wilk, Sz., 2000. Computer-assisted triage of abdominal pain in childhood: a suggested clinical algorithm. In: Proceedings of the Conference on Simulation in Health Sciences, San Diego, CA, Society for Computer Simulation 19-24.

20. Shapley, L.S., 1953. A value for n-person games. In: Kuhn, H.W., Tucker, A.W. (Eds.), Contributions to the Theory of Games II, Princeton University Press, Princeton, pp. 307-317.

21. Simpson, E.T., Smith, A., 1996. The management of acute abdominal pain in children, Journal of Paediatrics and Child Health 32 110-112.

22. Słowiński, R., 1995. Rough sets approach to decision analysis, AI Expert Magazine 10 8-25.

23. Stefanowski, J., Tsoukias, A., 2001. Incomplete information tables and rough classification, International Journal of Computational Intelligence 17 (3) 454-466.

24. Stefanowski, J., Vanderpooten, D., 2001. Induction of decision rules in classification and discovery-oriented perspectives, International Journal of Intelligent Systems 16 (1) 13-28.

25. Stefanowski, J., Wilk, Sz., 2001. Evaluating business credit risk by means of approach integrating decision rules and case based reasoning, International Journal of Intelligent Systems in Accounting, Finance and Management 10 97-114.

26. Ziarko, W., 1993. Variable precision rough sets model, Journal of Computer and Systems Sciences 46 (1), 39-59.