APPLICATION OF ROUGH SETS TO PATIENT SATISFACTION ANALYSIS

Yuerong Chen, University of Akron, USA, yuerong@uakron.edu
Shengyong Wang, University of Akron, USA, wangs@uakron.edu
Chien-Chung Chan, University of Akron, USA, chan@uakron.edu

ABSTRACT

Patient satisfaction survey has been widely used by healthcare providers to gauge their service level and initiate process improvement actions. Usually, summary statistics from the survey results such as means, percentile ranks, and correlation coefficients are analyzed for operational decisions. This paper presents a rough set approach to analyze patient satisfaction survey data collected from a medium-sized community hospital in USA. Advantages of this approach lie in its ability to process data with vagueness and uncertainty, data reduction capability, and strong interpretation ability of decision rules, which can be used to build rule-based inference systems.

Keywords *Patient satisfaction survey analysis, rough set theory, data reduction, decision rules*

INTRODUCTION

Hospitals are facing tremendous challenges to make the healthcare delivery system safe, effective, patient-centered, timely, efficient, and equitable. In pursuit of operational excellence, patient satisfaction survey has been widely used by the hospitals to measure their service levels during the course of healthcare delivery. According to the U.S. Department of Health and Human Services [1], these surveys are designed to provide hospitals with the information and insight on their service levels from patients' perspectives. Investigation of survey responses will assist hospitals to detect weakness in their services, develop strategies for service quality improvement, track service quality improvement over time, and compare their services with other competitors as well. There have been companies that provide exclusive service for patient satisfaction survey analysis. Representative of them is the notable Press Ganey Associates Corporate [2], which designs a set of standard survey questions for its customers (i.e., hospitals). The Press Ganey survey measures patients' perception of how well their experiences with a certain hospital were, using a rating scale of Very Good, Good, Fair, Poor, and Very Poor. The survey is composed of several sections with each section consisting of several questions to measure a specific aspect of the patient care received, such as patient care related to admission, room, nurses, physicians, etc. Patients rate their satisfaction level in each section. They also rate their perception with the overall care the hospital provides, as well as their likelihood to recommend this hospital to others.

Commonly, after the completed surveys are collected, survey responses are analyzed. Note that before conducting analyses, the responses need to be transformed into being of numerical form if the original rating scale is non-numerical. For example, the rating scale in the Press Ganey survey questionnaire can be transformed into {5, 4, 3, 2, 1}, with 5 standing for 'Very Good', 4 standing for 'Good', etc. The commonly used analysis methods for patient satisfaction survey include calculating means, standard deviations, correlation coefficients, percentile ranks, and so on. While correlation coefficient investigates the relationship between two survey questions, other analyses like mean and standard deviation look into a specific survey question only. For instance, percentile ranking tells a hospital of which percentile its score is compared with other hospitals with similar settings for the corresponding survey question. Obviously, the higher the percentile, the better the service of the concerned hospital compared to others. This method well demonstrates the competitiveness of a hospital.

In this paper, an approach using rough set theory is proposed for a more advanced analysis of patient satisfaction survey data. The rough set theory, introduced by Pawlak [3], has been proved to be an effective tool for handling the data with uncertainty or vagueness [4, 5]. This theory is of fundamental importance in artificial intelligence and cognitive sciences, especially in such research areas as machine learning, intelligent systems, inductive reasoning, pattern recognition, knowledge discovery, and expert systems [4]. The rough set approach has been applied in many fields, including bioinformatics [6, 7], medicine [8-10], business failure prediction [11, 12], computer network systems [13], business aviation decision making [14], acoustics [15], and other domains [16, 17]. For more applications, the readers are referred to [4, 18, 19]. In this paper, the rough set approach was used to reduce data dimensionality, i.e., to remove redundant survey questions and identify key survey questions that influence patient satisfaction, and to induce rules for investigating the methods to improve patients' overall satisfaction with the service/care a hospital provides.

The remainder of the paper is organized as follows. Section 2 presents the basic knowledge of the rough set theory. Section 3 describes the survey questionnaire under study and the preliminary

handling of the collected responses. Section 4 analyzes the survey data using rough sets and interprets the derived rules. Concluding remarks are included in Section 5.

BASICS OF THE ROUGH SET THEORY

Pawlak [3] introduced the rough set theory, which is a mathematical approach to tackling imperfect data (e.g., data with uncertainty or incompleteness). This approach is complementary to other uncertainty data handling methods such as probability theory, Dempster-Shafer theory of evidence, fuzzy set theory, and so on. The main advantage of the rough set theory lies in that it does not need any preliminary information about data like probability distributions in statistics, basic probability assignments in Dempster-Shafer theory, a grade of membership or the value of possibility in fuzzy set theory [4, 20-22]. Other advantages of the rough set approach can be found in [11, 23, 24].

Information Systems

An information system is defined as a 4-tuple $S = (U, Q, V, f)$, where U is a finite set of objects, Q is a finite set of attributes, $V = \cup_{q \in Q} V_q$ is a set of possible attribute values with V_q being the domain of the attribute q, and $f: U \times Q \to V$ is a function such that $f(x, q) \in V_q$ for every $x \in U, q \in Q$.

The mathematical basis of the rough set theory is indiscernibility relation. Let $B \subseteq Q$ and $x, y \in U$, x and y are indiscernible with respect to B if and only if $f(x, q) = f(y, q)$ for every $q \in B$. Given a certain $B \subseteq Q$, a set of indiscernible objects comprise an equivalence class. Obviously, different $B \subseteq Q$ has different equivalence classes. Equivalence classes of B are also called B-elementary sets. We denote by $I_B(x)$ a B-elementary set that contains the object $x \in U$.

Approximations and Quality of Classification

Lower approximation and upper approximation are two basic operations in rough set theory. Let $B \subseteq Q$ and $X \subseteq U$. The B-lower approximation and B-upper approximation of X, denoted by $\underline{B}(X)$ and $\overline{B}(X)$, respectively, are defined as follows:

$$\underline{B}(X) = \{x \in U: I_B(x) \subseteq X\}, \quad (1)$$

$$\overline{B}(X) = \{x \in U: I_B(x) \cap X \neq \emptyset\}. \quad (2)$$

The difference between the two approximations is defined as the B-boundary region of X, which is denoted by $BN_B(X)$ and calculated by the following equation:

$$BN_B(X) = \overline{B}(X) - \underline{B}(X). \quad (3)$$

$\underline{B}(X)$ is a set of objects that can be certainly classified as the elements of X in terms of the set of attribute B, $\overline{B}(X)$ is a set of objects that can be possibly classified as the elements of X in terms of B, and $BN_B(X)$ is a set of objects that cannot be certainly classified to X.

If $BN_B(X) = \emptyset$, i.e., $\underline{B}(X) = \overline{B}(X)$, the set X is said to be crisp or exact with respect to B. On the other hand, if $BN_B(X) \neq \emptyset$, then the set X is referred to as rough or inexact with respect to B. That is, a rough set, in contrast to a crisp set, has a non-empty boundary region. Any rough set is associated with the accuracy of approximation by B, which is defined as

$$\alpha_B(X) = \frac{card(\underline{B}(X))}{card(\overline{B}(X))} \quad (4)$$

where $card(\cdot)$ denotes the cardinality of a set. Clearly, $0 \le \alpha_B(X) \le 1$. If $\alpha_B(X) = 1$, then X is crisp with respect to B.

Another important concept in rough set theory is the quality of approximation of classification, or in short, quality of classification. Let $\mathcal{Y} = \{Y_1, Y_2, \cdots, Y_k\}$ be a mutually exclusive classification (or partition) of the universe U. The quality of classification, denoted by $\gamma_B(\mathcal{Y})$, is defined as below:

$$\gamma_B(\mathcal{Y}) = \frac{\sum_{i=1}^{k} card(\underline{B}(Y_i))}{card(U)} \quad (5)$$

It expresses the ratio of all B-correctly classified objects to all objects in the system. If $\gamma_B(\mathcal{Y}) = 1$, then the classification $\mathcal{Y}$ is crisp (or precise) with respect to the set of attributes B.

Reduct and Core

One of the most important contributions of rough set theory to the data analysis field is that it can remove

superfluous information. The involved concepts are reduct and core. A reduct $R \subseteq Q$ of an information system S is defined as a minimal subset of attributes that preserves the quality of classification as the set of all attributes Q. That is, $\gamma_R(U) = \gamma_Q(U)$. Usually, there exist more than one reduct for an information system. The intersection of all reducts is called a core. The core is therefore the most important subset of attributes, since none of its elements can be removed without affecting the classification power of attributes.

Decision Rules

Sometimes the set of attributes Q can be divided into two mutually exclusive sets C and D (i.e., $Q = C \cup D, C \cap D = \emptyset$), with C containing condition attributes and D containing decision attributes. Such an information system $S = (U, C \cup D, V, f)$ is also called a decision system. A set of decision rules can be derived from a decision system.

A decision rule is any expression of the form $\Phi \rightarrow \Psi$, which reads that if Φ holds, then Ψ holds. Φ is a conjunction of elementary condition formulae, i.e., $(a_1 = v_1) \wedge \cdots \wedge (a_s = v_s)$, while Ψ is a disjunction of elementary decision formulae, i.e., $(d = k_1) \vee \cdots \vee (d = k_t)$. If $\tau = 1$, the decision rule is exact. Otherwise, it is approximate. An approximate rule means that, using the available knowledge, some objects cannot be definitely assigned to a decision class. An important characteristic of a decision rule is its support, which is defined as the number of the objects that obey this rule. Several procedures have been proposed [25-[28] for inducing decision rules from a decision system.

Decision Support or Classification

The derived decision rules can be used for decision support of a new object (i.e., classifying a new object into a decision class). Specifically, this is done by matching the description (or the condition attribute values) of the new object to one of the decision rules. Accordingly, there are four possible situations [29, 30]:

(1) If the new object matches a single exact rule, then the new object is classified into the decision class pointed by the rule.
(2) If the new object matches more than one exact rule that indicate, however, the same decision class, then the object is assigned to that decision class.
(3) If the new object matches one approximate rule or several exact rules that indicate different decision classes, then the object is classified into the decision class which has the largest support.
(4) If the new object does not match any of the rules, then the object is assigned to a decision class by the nearest rule, i.e., nearest according to the selected distance metric (the Lp-metric or valued closeness relation [29]). If there are several rules with the minimal distance to the object, then the object is assigned to the class with the largest support.

DESCRIPTION OF PATIENT SATISFACTION SURVEY DATA

The concerned patient satisfaction survey was conducted in a community hospital in the United States. The survey population are the inpatients who were discharged during the period from October 2008 to September 2009. Press Ganey survey questionnaires were used, including 37 standard questions which are listed in Appendix. These questions cover 10 sections: admission section (a), room section (r), food section (m), nurse section (n), test section (t), visitor section (v), physician section (p), discharge section (d), staff section (i), and overall care section (o). The rating scale for these questions is {Very Good, Good, Fair, Poor, and Very Poor}. For the convenience of analysis, this scale is converted into a numerical form {5, 4, 3, 2, 1}, with 5 standing for 'Very Good', 4 standing for 'Good', etc.

The hospital received 815 responses totally, among which 507 responses answered all 37 standard questions. These 507 survey responses comprise the information system under study. For simplification, the responses for the 3 questions in the overall care section are averaged, and the rounded average is taken as a single decision attribute which is denoted by the label 'AvgO'. The other 34 questions are considered to be condition attributes. Among the 507 complete responses, the number and percentage of the responses that have a given value of 'AvgO' are shown in Table 1. Most patients considered the overall service of this hospital 'Very Good' (73.37%); a considerable number of patients rated it 'Good' (20.32%); only a few patients felt the service 'Fair', 'Poor', and 'Very Poor' (1.18%, 0.99%, and 4.14%, respectively). Table 1 reflects that, the service/care this hospital provided in the concerned period is very good from the standpoint of patients.

Table 1 The number and percentage of the responses for each value of 'AvgO'

AvgO	1	2	3	4	5
Number of Responses	6	5	21	103	372

Percentage of Responses (%)	1.18	0.99	4.14	20.32	73.37

The significant differences among the numbers of the responses for these 5 satisfaction levels make it difficult to conduct accurate classification analysis of patient satisfaction survey. Therefore, combined with the consideration of the structure of the responses for 'AvgO', the following transformation is implemented on the information system: the satisfaction levels {1, 2, 3} are regarded as a single level and denoted by '3'. In other words, the entries of {1, 2} in the investigated information system are changed into '3'. Thus, the adjusted information system includes 3 levels only, representing 'Not Good', 'Good', and 'Very Good', respectively. As Table 1 indicates that most patients were very satisfied with the service/care they received in this hospital, the analysis of the transformed information system is expected to lead to the following results: (1) to discover what aspects of the service the patients considered the most unacceptable so that they gave the average overall satisfaction score of '3'; (2) to explore what aspects of the service can be improved so that the average overall satisfaction score can change from '4' to '5'; (3) to find out what aspects of the service are critical for the hospital to receive a score of '5' on its overall service/care.

SURVEY ANALYSIS AND RESULTS

The rough set analysis was conducted on the transformed information system, and was performed using the software ROSE2 which was created at the Laboratory of Intelligent Decision Support Systems of the Institute of Computing Science in Poznan [30]. The analysis of the transformed information system shows the following results:

(1) With respect to the set of all condition attributes (i.e., 34 standard questions in the sections other than the overall care section), the approximation of each decision class is perfect. In other words, $\alpha_Q(Y_i) = 1, i = 1,2,3.$ where Q is the set of all condition attributes and $Y_i (i = 1,2,3)$ denotes each decision class, with the value of 'AvgO' being 3, 4, and 5, respectively. Besides, the quality of classification is perfect as well, namely, $\gamma_Q(\mathcal{Y}) = 1$, where $\mathcal{Y} = \{Y_1, Y_2, Y_3\}$ is the classification of all 507 responses based on the value of the decision attribute of 'AvgO'. The perfect approximation of each decision class and the perfect quality of classification imply that all responses in the information system (patient satisfaction survey results) can be clearly distinguished into each decision class, i.e., the classes of 'Not Good', 'Good', and 'Very Good'.

(2) The core of the attributes is empty. According to the definition of core in Section 2.3, this means that no single attribute is absolutely necessary for the perfect approximation of each decision class and the perfect quality of classification. If there is a non-empty core, then the attributes contained in the core are the most important and indispensable attributes for preserving the quality of classification. This implies among the 34 survey questions, not a single survey question can be identified as the most significant question affecting patient overall satisfaction.

(3) There are a considerable number of reducts derived from the information system under study. This is due to the relatively large number of condition attributes (i.e., 34) and the significant unbalance among the decision classes. As far as the shortest reducts are concerned, there are totally 138 such reducts, with each including 9 attributes. Based on the selection criteria adopted in [11, 14], the most significant attributes from the perspective of the decision makers (the hospital managers) are included and the number of attributes are kept as small as possible. The following reduct was selected for further analysis: {a1 (Speed of admission), r2 (Room cleanliness), r4 (Room temperature), r5 (Noise level in and around room), m3 (Quality of the food), n2 (Promptness response to call), t5 (Courtesy of person took blood), p5 (Skill of physician), d2 (Speed of discharge process)}. This dramatically reduces the size of the information system, while keeping the quality of classification according to the definition of a reduct. In other words, the above 9 survey questions are identified as the most significant questions that impact the patient satisfaction rating.

The next step is to study how the different combinations of the responses to the 9 survey questions lead to different patient overall satisfaction ratings, based upon which a set of "decision rules" will be generalized. This is essentially useful for hospital managers to quantitatively decide what aspects of the care delivery process need to be improved. Reserving the 9 attributes in the selected reduct and removing the other 25 attributes from the information system result in a reduced information table, from which a set of decision rules will be derived. A rule induction algorithm LEM2 (Learning from Examples Module, Version 2) [25] was used to generate a minimal set of rules covering all objects in the reduced information system.

There are a total of 99 rules derived. Table 2 shows part of the rules for each decision class. For the

decision class of $AvgO = 3$, only the rules with support greater than or equal to 5 are considered. For the other two decision classes, the requirements of support greater than or equal to 7 and support greater than or equal to 10 are executed, respectively. Such actions are based on the consideration of different numbers of responses in the decision classes. Other rules that do not satisfy the above support requirements are regarded too weak to take into account. The rules in the tables are easy to understand. Take Rule 1 in Table 2 as an example. It means that, if a patient rates $r2$ (Room cleanliness) as "Not good", and $p5$ (Skill of physician) as "Not good", then the patient's overall satisfaction rating is "Not good", regardless of the patient's other ratings to other questions. Moreover, there are 8 survey responses supporting this rule. The shorter a rule, the stronger it is. Also, the larger the support of a rule, the stronger the rule. The accuracy rate of the derived decision rules is also evaluated. The average classification accuracy rate is 81.07%, based on 10 times of 10-fold cross validation.

Table 2 The decision rules for the decision class of AvgO = 3, with support >= 5

Rule #	a1	r2	r4	r5	m3	n2	t5	p5	d2	AvgO	Sup
1		3						3		3	8
2		3			3	3			3	3	6
3			3	3				3		3	6
4		3		4		3			3	3	5

Table 2 contains the decision rules for the decision class of $AvgO = 3$ with support >= 5. Rule #1 in Table 2, as discussed above, is a strong rule, since its length is only 2. It reflects that, the hospital should pay attention to these two questions $r2$ (room cleanliness) and $p5$ (skill of physician), because that if both aspects are dissatisfying (i.e., patients rate them 'Not Good'), the hospital will receive a dissatisfactory rating about its overall service/care. Besides, it is found from Table 2 that $r2 = 3$ (room cleanliness rated as "Not good") appears three times in the above 4 rules. This implies that, if a patient is not satisfied with room cleanliness, then the chance of the hospital receiving a dissatisfactory overall rating is high. Therefore, the hospital should try to improve $r2$'s rating (i.e., make rooms clean to make patients comfortable and happy). Similarly, since $n2$ (nurses' promptness of response to call), $p5$ (skill of physician), and $d2$ (speed of discharge process) receive low scores twice in the above rules, they should also be noted. In Rule #4 in Table 2, $r5 = 4$ ("Noise level in and around room" rated as "Good"), however, the overall rating is still not good (i.e., $AvgO = 3$). This indicates that, the existence of good service in few aspects does not change too much patients' impression of the hospital, if other aspects perform below average.

Table 3 The decision rules for the decision class of AvgO = 4, with support >= 7

Rule #	a1	r2	r4	r5	m3	n2	t5	p5	d2	AvgO	Sup
1				4		4	4	4	4	4	20
2				4	4		4	4	4	4	20
3			4	4	4	4		4	4	4	11
4		4		4	4	3				4	9
5		4	4			3			4	4	9
6	5	4	4			4	4	4		4	7
7	5			4	4			4	4	4	7

Table 4 The decision rules for the decision class of AvgO = 5, with support >= 10

Rule #	a1	r2	r4	r5	m3	n2	t5	p5	d2	AvgO	Sup
1			5			5				5	158
2		5				5	5		5	5	150
3		5	5					5		5	138
4			5	5						5	135
5				5	5			5		5	115
6	5		4			5		5		5	71
7		4		5				5		5	26
8		5	4	4			5	5		5	25
9		5		4	5				5	5	23
10			4	4	4	5			5	5	22
11					5	5			4	5	21
12	4						5		5	5	20
13	5						5	4	5	5	17
14		5	4	5		4				5	10

Table 3 shows the decision rules for the decision class of $AvgO = 4$ with support >= 7. It can be seen from this table that, $AvgO = 4$ when the number of the survey questions whose score is 4 is larger than or equal to 4. It makes sense since the overall rating is based on the perception of patients on each aspect of the service. It is also observed that, in all rules of Table 3, the following questions have higher frequency of receiving a rating of 'Good' (i.e., $AvgO = 4$): $r5$ (noise level in and around room), $m3$ (quality of the food), $p5$ (skill of physician), and $d2$ (speed of discharge process). This demonstrates that, these questions are critical for the hospital to achieve a good rating of its overall service/care. Paying more attention to these aspects may make patients more satisfied and lead to a better overall rating. In Rules #4 and #5, $n2 = 3$, however, the hospital still receives an overall rating of 'Good'.

Since n2 denotes the nurses' promptness of response to call, this may reflect that, the response of the nurses to call is not quick and this aspect should be emphasized and improved. It also means that, bad performance of few aspects does not influence the overall rating too much, if other aspects perform well. The patients have tolerance for some aspects. On the other hand, a1 5 in Rules #6 and #7. It shows the opposite. In other words, good performance of few aspects does not have a big impact the overall rating.

Table 4 is a set of decision rules for the decision class of AvgO 5 with support >= 10. Rules #1 and #4 are strong rules with the length of 2 and big support. They explore that, r4 (room temperature) and n2 (nurses' promptness of response to call) (or r5 (noise level in and around room)) are significant survey questions for patients to think of the overall service/care of the hospital 'Very Good'. This is because that, if both r4 and n2 (likewise, r4 and r5) score 5, then the hospital receives an overall rating of 'Very Good'. It may also indicate that, r4 and n2 (likewise, r4 and r5) are the aspects that usually perform worse than the other aspects do, therefore, if they are greatly improved to make patients very satisfactory, then the overall rating can reach '5'. Thus, if the hospital wants to have a 'Very Good' overall service/care, it should pay attention to r4 and n2 (likewise, r4 and r5). Furthermore, in all rules in Table 4, the frequency of r4 4 and that of r5 4 are high. This further verifies that, r4 and r5 are the aspects that have worse performance compared with other aspects in this class. Although patients give an overall rating of 'Very Good', they are apparently tolerating with these two aspects. Therefore, the hospital should start to come up with ideas to improve the service in these two aspects. In addition, the other aspects that need special notice include r2 (room cleanliness), p5 (skill of physician), and d2 (speed of discharge process), since their occurrence rates in Table 4 are high. In order to receive a highest score for its overall service/care, the hospital ought to make the performance of these aspects outstanding. This is because that the chance of achieving a 'Very Good' overall service greatly increases if these aspects perform very well.

In sum, to avoid having a bad overall rating of '3', the hospital should pay close attention to the following aspects: r2 (room cleanliness), n2 (nurses' promptness of response to call), p5 (skill of physician), and d2 (speed of discharge process). For receiving a good rating of its overall service/care, the following aspects should be highly concerned: r5 (noise level in and around room), m3 (quality of the food), n2 (nurses' promptness of response to call), p5 (skill of physician), and d2 (speed of discharge process). Lastly, for achieving a very good overall rating, the hospital needs to make sure of very good performance of the following aspects: r2 (room cleanliness), r4 (room temperature), r5 (noise level in and around room), n2 (nurses' promptness of response to call), p5 (skill of physician), and d2 (speed of discharge process). It is noticed that, n2 (nurses' promptness of response to call), p5 (skill of physician), and d2 (speed of discharge process) all appear in the above three suggestions. It implies that, these three aspects are very critical and the hospital should try to make the performance of these aspects very good. Moreover, it is noted that, a1 (speed of admission) and r5 (courtesy of person took blood) do not appear in the above three suggestions. This indicates that, these two aspects do not have big influences on the overall rating of the hospital.

CONCLUDING REMARKS

Patients' satisfaction is a very important measure of the care/service provided by hospitals. Almost every hospital in the United States conducts patient satisfaction survey to evaluate their service performance. Common statistical approaches, such as percentile ranks, correlation analysis, have been used to analyze to survey data for hospital operational decisions. In this paper, a novel approach using the rough set theory was studied to analyze the patient satisfaction survey data obtained from a community hospital in USA. The rough set approach has many advantages over other approaches, especially in its ability of dealing with vague data, reducing data size, and inducing a set of decision rules that are easy to interpret. In this paper, 9 out of 34 survey questions are first identified as the most significant questions that influence patient overall satisfaction ratings. A set of useful decision rules are then derived to provide the hospital with some insightful suggestions on how to improve its service. Specifically, the aspects of the service that need to be paid close attention to are pointed out. Overall, the rough set

approach is proven to be a useful tool for the analysis of patient satisfaction survey.

Appendix

Explanations to Standard Questions in Press Ganey Patient Satisfaction Survey

Question	*Description*
a1	Speed of admission
a2	Courtesy of person admitting
r1	Pleasantness of room decor
r2	Room cleanliness
r3	Courtesy of person cleaning room
r4	Room temperature
r5	Noise level in and around room
m2	Temperature of the food
m3	Quality of the food
m4	Courtesy of person served food
n1	Friendliness/courtesy of the nurses
n2	Promptness response to call
n3	Nurses' attitude toward requests
n4	Attention to special/personal needs
n5	Nurses kept you informed
n6	Skill of the nurses
t1	Wait time for test or treatments
t3	Explanations: happen during T&T
t5	Courtesy of person took blood
t7	Courtesy of person started IV
v2	Accommodations & comfort visitors
v3	Staff attitude toward visitors
p1	Time physician spent with you
p2	Physician concern questions/worries
p3	Physician kept you informed
p4	Friendliness/courtesy of physician
p5	Skill of physician
d1	Extent felt ready discharge
d2	Speed of discharge process
d3	Instructions care at home
i1	Staff concern for your privacy
i3	How well your pain was controlled
i4	Staff addressed emotional needs
i5	Response concerns/complaints
i6	Staff include decisions re:trtmnt
o2	Staff worked together care for you
o3	Likelihood recommending hospital
o4	Overall rating of care given

REFERENCES

[1] http://bphc.hrsa.gov/patientsurvey/

[2] http://www.pressganey.com/

[3] Pawlak Z (1982) Rough sets. International Journal of Information and Computer Sciences 11: 341-356

[4] Pawlak Z, Skowron A (2007) Rudiments of rough sets. Information Sciences 177: 3-27

[5] Pawlak Z (1991) Rough sets – theoretical aspects of reasoning about data. Kluwer Academic Publishers, Dordrecht, Boston, London

[6] Midelfart H, Komorowski J, Nørsett K, Yadetie F, Sandvik AK, Lægreid A (2004) Learning rough set classifiers from gene expression and clinical data. Fundamenta Informaticae 2 : 155-183

[7] Hvidsten TR, Wilczyński B, Kryshtafovych A, Tiuryn J, Komorowski J, Fidelis K (2005) Discovering regulatory binding-site modules using rule-based learning. Genome Research 15: 856-866

[8] Bazan J, Osmólski A, Skowron A, Ślezak D, Szczuka M, Wróblewski J (2002) Rough set approach to the survival analysis. In: J.J. Alpigini et al. (Eds.): RSCTC 2002, LNAI 2475, pp. 522-529

[9] Grzymaa-Busse JW, Hippe ZS (2005) Data mining methods supporting diagnosis of melanoma, In: 18th IEEE Symposium on Computer-Based Medical Systems (CBMS 2005), 23–24 June 2005, Dublin, Ireland, IEEE Computer Society, pp. 371–373

[10] Hirano S, Tsumoto S (2005) Rough representation of a region of interest in medical images. International Journal of Approximate Reasoning 40: 23-34

[11] Dimitras AI, Słowiński R, Susmaga R, Zopounidis C (1999) Business failure prediction using rough sets. European Journal of Operational Research 114: 263-280

[12] Beynon MJ, Peel MJ (2001) Variable precision rough set theory and data discretisation: an application to corporate failure prediction. Omega 29: 561-576

[13] Cai Z, Guan X, Shao P, Peng Q, Sun G (2003) A rough set theory based method for anomaly intrusion detection in computer network systems. Expert Systems 20: 251-259

[14] Ou Yang YP, Shieh HM, Tzeng GH, Yen L, Chan CC (2010) Combined rough sets with flow graph and formal concept analysis for business aviation decision-making. Journal of Intelligent Information Systems.

[15] Kostek B (1999) Soft computing in acoustics, applications of neural networks, fuzzy logic and rough sets to musical acoustics. Studies in Fuzziness and Soft Computing. New York: Heidelberg

[16] Zhai LY, Khoo LP, Fok SC (2002) Feature extraction using rough set theory and genetic algorithms – an application for the simplification of product quality evaluation. Computers & Industrial Engineering 43: 661-676

[17] Lee S, Vachtsevanos G (2002) An application of rough set theory to defect detection of automotive glass. Mathematics and Computers in Simulation 60: 225-231

[18] Słowiński R (1992) Intelligent decision support – handbook of applications and advances of the rough sets theory. Kluwer Academic Publishers, Dordrecht

[19] Pawlak Z (1997) Rough set approach to knowledge-based decision support. European Journal of Operational Research 99: 48-57

[20] Dubois D, Prade H (1992) Putting rough sets and fuzzy sets together. In: Słowiński R (Ed.), Intelligent Decision Support: Handbook of

Applications and Advances of the Rough Sets Theory. Kluwer Academic Publishers, Dordrecht, 203-232

[21] Krusińska E, Słowiński R, Stefanowski J (1992) Discriminant versus rough set approach to vague data analysis. Applied Stochastic Models and Data Analysis 8: 43-56

[22] Skowron A, Grzymaa-Busse JW (1993) From the rough set theory to the evidence theory. In: Fedrizzi M, Kacprzyk J, Yager RR (Eds.), Advances in the Dempster-Shafer Theory of Evidence. Wiley, New York, 295-305

[23] Greco S, Matarazzo B, Słowiński R (1998) A new rough set approach to evaluation of bankruptcy risk. In: Zopounidis C (Ed.), Operational Tools in the Management of Financial Risks. Kluwer Academic Publishers, Dordrecht, 121-136

[24] Pawlak Z (2002) Rough sets and intelligent data analysis. Information Sciences 147: 1-12

[25] Grzymaa-Busse JW (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 Sets Theory. Kluwer Academic Publishers, Dordrecht, 3-18

[26] Skowron A (1993) Boolean reasoning for decision rules generation. In: Komorowski J, Ras ZW (Eds), Methodologies for Intelligent Systems, Lecture Notes in Artificial Intelligence, vol. 689. Springer, Berlin, 295-305

[27] Stefanowski J, Vanderpooten D (1994) A general two-stage approach to including rules from examples. In: Ziarko W (Ed.), Rough Sets, Fuzzy Sets and Knowledge Discovery. Springer, London, 317-325

[28] Stefanowski J (1998) On rough set based approaches to induction of decision rules. In: Skowron A, Polkowski L (Eds.), Rough Sets in Knowledge Discovery, vol. 1, Physica Verlag, Herdelberg, 500-529

[29] Słowiński R, Stefanowski J (1994) Rough classification with valued closeness relation. In: Diday E, et al. (Eds.), New Approaches in Classification and Data Analysis. Springer, Berlin, 482-488

[30] Prędki B, Wilk S (1999) Rough set based data exploration using ROSE system. In: Ras ZW, Skowron A (Eds.), Foundations of Intelligent, Lecture Notes in Artificial Intelligence, vol. 1609, Springer, Berlin, 172-180