In Hoffman F, Hand DJ, Adams N, Fisher D & Guimaraes G (Eds.) Lecture Notes in Computer Science 2189: Advances in Intelligent Data Analysis, Fourth International Conference (IDA-01),2001, Cascais Portugal. Springer Verlag: Berlin.

Analyzing Data Clusters: A Rough Set Approach to Extract Cluster-Defining Symbolic Rules

Syed Sibte Raza Abidi, Kok Meng Hoe, Alwyn Goh

School of Computer Science, Universiti Sains Malaysia, 11800 Penang, Malaysia
{sraza, kmhoe, alwyn}@cs.usm.my

Abstract. In this paper we present a strategy together with its computational implementation to intelligently analyze data clusters in terms of symbolic cluster-defining rules. We present a symbolic rule extraction workbench that leverages *rough set* theory to inductively extract CNF form symbolic rules from un-annotated continuous-valued data-vectors. Our workbench purports a hybrid rule extraction methodology, incorporating a sequence of methods to achieve data clustering, data discretization and eventually symbolic rule discovery via rough set approximation. The featured symbolic rule extraction workbench will be tested and analyzed using several well-known biomedical datasets.

1. Introduction

The on-going *information revolution* is generating volumes of data, from sources as diverse as banking transactions, scientific explorations, telecommunication networks, space science, medical systems, human genome research and so on. Indeed, there is an imperative on the intelligent analysis of such large volumes of data so as to derive intrinsic strategic knowledge—knowledge encoded in terms of trends, patterns, associations, constraints, business rules, etc.—that can impact to optimize decision-support, business competitiveness and other services-oriented portfolios.

Data clustering is a popular data analysis task that involves the distribution of 'un-annotated' data (i.e. with no a priori class information), in an inductive manner, into a finite sets of categories or clusters such that data items within a cluster are similar in some respect and unlike those from other clusters. If one regards data as an underlying quantitative statement about a system's behavior—either human or engineered—within a particular environment, then exploratory data clustering algorithms attempt to learn the topology of the data by analyzing the inherent similarities and differences of the individual data items in the untagged data set.

Notwithstanding the efficacy of traditional data clustering techniques, it can be argued that the outcome of a data clustering task does not necessarily explicates the intrinsic relationship between the various attributes of the dataset. What we mean here is that the output of a data clustering task does not provide the value-added knowledge—most preferably in a symbolic formalism such as deductive rules—defining both the structure of the emerged clusters and the cluster membership principles. From an intelligent data analysis perspective, cluster-defining knowledge

is highly desirable as it can provide interesting insights into the complex inter-relationships between the various data attributes.

The featured work is motivated by the desirability of deriving cluster-defining knowledge for a priori defined data clusters [1]. We present a multi-strategy approach for the automated extraction of cluster-defining *Conjunctive Normal Form* (CNF) symbolic rules from un-annotated data-sets. The motivation for our work stems from the individual effectiveness of various data analysis mechanisms: (1) cluster formation via unsupervised clustering algorithms, (2) data-set simplification and attribute selection via attribute discretization, and (3) symbolic rule extraction via rough set approximation [2]. We have implemented a generic *Symbolic Rule Extraction Workbench* (see Figure 1) that can generate cluster-defining symbolic rules from continuous-valued data, such that the emergent rules are directly applicable to rule-based expert systems [3].

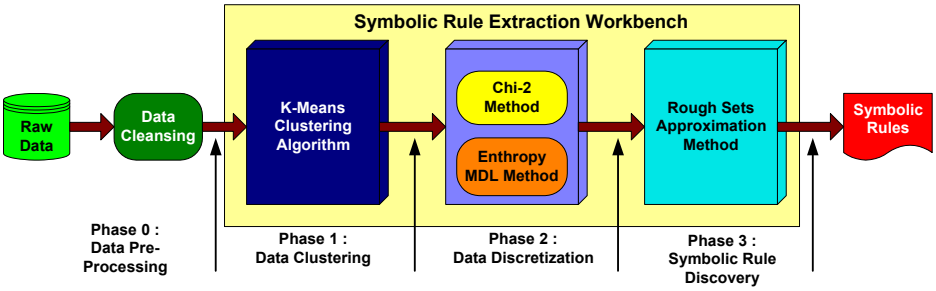

Fig. 1. The Functional Architecture of the Symbolic Rule Extraction Workbench

2. Rough Sets: A Brief Overview

The main objective of rough set mediated data analysis is to form approximate concepts about the data based on available classification or decision information [2]. These data-defining approximate concepts generated via rough set analysis are typically represented as succinct symbolic rules that provide an explanation about inter-attribute dependencies, attribute importance and topology-defining information vis-à-vis an annotated data-set.

In rough set framework, annotated data is represented as a decision system, Γ defined as a pair $\Gamma = (U,A)$, where U is a finite set of objects called the *universe* and A is a finite set of *attributes*. For every $a \in A$, $a{:}U \rightarrow V_a$, where $V_a \neq \varnothing$ is called the *values set* of a. Attributes in A are divided into two disjoint sets, $A = C \cup \{d\}$, where $\{d\} \notin C$ is the singular decision or class attribute and C is the set of condition attributes. Hence Γ can be denoted as $\Lambda = (U, C \cup \{d\})$.

Rough sets based data analysis, leading to symbolic rule extraction involves the following processing steps:

- Definition of an approximate space by finding *indiscernibility relations* between objects in the universe. Two objects, x, $y \in U$ are indiscernible when they are equivalent with regards to their attributes and values. With any subset $B \subseteq A - \{d\}$

or B $\subseteq$ *C*, an *indiscernibility relation I(B)* partitions the universe into separate sets called *equivalence classes*, $[x]_{I(B)}$, which denotes the set of all objects equivalent to *x* in terms of all attributes in *B*. The indiscernibility relations are used to reduce the size of the universe such that only a single element of the equivalence class is required to represent the characteristics of the entire class.

- Concept approximation is achieved by the reduction of the data-set by retaining only those attributes that contribute towards the preservation of the indiscernibility relation. The minimum subset of attributes retained to maintain 'indiscernibility' of all objects in the data-set—i.e. the minimal set of attributes that can differentiate all equivalence classes in the universe—are called *reducts*. A set of reducts are next computed by eliminating superfluous attributes *a* when there exist $B' \subseteq B$ where $a \in B$, $a \notin B'$ and $I(B') = I(B)$.
- Symbolic decision rules can next be synthesized from the set of reducts. A decision rule provides a definitive description of the concepts within the universe in terms of a statement of the form "if *Conditions* are True then *Outcomes* are True". In practice, a decision rule is synthesized by superimposing the reduct with the decision system—i.e. by taking for each attribute in the reduct its corresponding value from objects in the dataset together with their decision values. For instance, given the reduct $\{a_1, a_3\}$, where $a_i \in C$, $i = 1,\ldots,|C|$ and the values set of a_i, V_a^i and the values set of *D*, V_d, where $d \in D$, the decision rule for Γ is

  IF $a_1 = V_a^1(j)$ and $a_3 = V_a^3(j)$ THEN $D = V_d(j)$, for $j = 1, 2, \ldots, |U|$. **(1)**

  Finally, the decision rules are assessed in terms of their measure of *accuracy*—i.e. how well they perform in predicting the class or outcome of new data patterns.

3. Extracting Cluster-Defining Symbolic Rules: A Multi-Strategy Approach

In our work, we intend to extract symbolic rules from un-annotated datasets comprising an undifferentiated collection of continuous-valued multi-component data-vectors $S = \{\mathbf{X}_i : \; i \in [1, n]\}$, for which the classification attribute $c(\mathbf{X}_i) = \alpha$ for $\alpha \in [1, k]$ is unknown [5]. We have postulated a multi-strategy approach that dictates the systematic transformation of un-annotated data-sets to deductive symbolic rule-sets via a sequence of phases, as described below:

3.1 Phase 1 - Data Clustering

Given an un-annotated dataset satisfying the above assumption, we first partition it into *k* clusters, where each cluster comprises data-vectors with similar inherent characteristics. Note that the data clustering task is carried out with no a priori knowledge about the intrinsic class structure—i.e. how the data is inherently partitioned into distinct clusters. In practice, the data clustering algorithm inductively

derives the class information and partitions the data-set accordingly. We use the popular *K-Means* data clustering algorithm primarily due to its effectiveness and procedural simplicity [6]. The net outcome of this phase is the availability of k number of data clusters, which forms the basis for subsequent discovery of symbolic rules that define the structure of the discovered clusters.

3.2 Phase 2 - Data Discretisation

The motivation for this phase is driven by the fact that ordinal or continuous valued attributes are proven to be rather unsuitable for the extraction of concise symbolic rules. Henceforth, the necessity to discretise continuous-valued attributes to discrete intervals—i.e. reduce the domain of values of an attribute to a small number of attribute-value ranges—where each interval can be represented by a label/token. More attractively, the data discretization phase not only reduces the complexity and volume of the data-set, but also serves as a attribute filtering mechanism, whereby attributes that are deemed to have minimum impact on the class specification can be eliminated. In our work, we employ two data discretisation methods: (1) statistical discretization via Chi-2 [7] and (2) class information entropy reduction via MDL partitioning [8]; their respective results provide for an interesting contrast.

3.3 Phase 3 - Symbolic Rule Discovery

We use rough set approximation—an interesting alternative to a variety of symbolic rule extraction methods [9,10]—to derive symbolic rules that explain the inherent dependencies, attribute significance and structural characteristics of the annotated and clustered data-set. We have devised a three step methodology for the generation of symbolic rules from annotated data.

3.3.1 Step 1: Construction of Dynamic Reducts

First we randomly partition the discretized data into two disjoint sets: a bigger *training set* (70% of the dataset) and a smaller *testing set* (30% of the dataset). Next, we create 50 sub-samples of the training data-set by selecting 10 different random samples comprising 90% of the training data, and likewise 10 different random samples each of 80%, 70%, 60% and 50% of the training data. The rationale for this approach is to give multiple perspectives of the training data to the rule discovery algorithm in order to accumulate a larger set of rules and also to ensure a self-critiquing mechanism, whereby inter-sample rules critique each other for rule veracity purposes. From the 50 sub-samples of the training data, we compute multiple dynamic reduct-sets using genetic algorithm based methods [11,12,13,14]. Finally, we select dynamic reducts—i.e. those reducts that have a high frequency of occurrence across all the available reduct-sets—from all the available reduct-sets generated from the multiple data sub-samples. Standard search methods are employed over all the available reduct-sets to collect a unified set of dynamic reducts that is representative of the entire data.

In Hoffman F, Hand DJ, Adams N, Fisher D & Guimaraes G (Eds.) Lecture Notes in Computer Science 2189: Advances in Intelligent Data Analysis, Fourth International Conference (IDA-01),2001, Cascais Portugal. Springer Verlag: Berlin.

3.3.2 Step 2: Generate Symbolic Rules

We proceed to generate symbolic rules from the set of dynamic reducts. Instead of using all the dynamic reducts to generate a large set of symbolic rules, we attempt to generate symbolic rules from the shortest possible length dynamic reducts; the rationale being that shorter length dynamic reducts have been shown to yield concise rule-sets that exhibit higher classification accuracy and generalization capabilities [4]. This implies the need to initially select dynamic reducts of the shortest possible length followed by the generation of rules from the selected dynamic reducts. Our rule generation strategy therefore involves: (1) the selection of dynamic reducts that have a short length and (2) the generation of rules that satisfy a user-defined accuracy level. Our strategy for generating symbolic rules is as follows:

Step 1 : Specify an acceptable minimum accuracy level for the rule set.

Step 2 : Generate 50 random sub-samples of different sizes as described in section (3.3.1).

Step 3 : Find dynamic reducts from the sub-samples and place in set *DR*. Note that *DR* will comprise reducts with varying lengths.

Step 4 : From the reducts in *DR* determine the shortest reduct length (*SRL*).

Step 5 : From *DR*, collect all reducts that have a length equal to *SRL* and store them as set *SHRED*.

Step 6 : Generate symbolic rules from the reducts placed in *SHRED*.

Step 7 : Determine the overall accuracy of the generated rules with respect to the test data.

Step 8 : IF Overall accuracy of the generated rules is lower than the minimum accuracy level AND there exist reducts in the *DR* set with length > *SRL*
THEN Empty *SHRED* AND Update the value of *SRL* to the next highest reduct length in *DR* AND Repeat from step 6.
ELSE
Symbolic rules with the desired accuracy level cannot be generated.

At the conclusion of the above rule generation strategy we will ideally have a non-empty SHRED that will contain reducts yielding a set of rules that satisfy the acceptable accuracy level.

3.3.3 Step 3: Rule Filtering

The rule-set generated in Step 2 (3.3.2) will then undergo a filtering process guided by the following filtering criteria: (1) Right-hand-side (RHS) support of a rule—i.e. the number of patterns in the training set which support the consequent of a rule; (2) Left-hand-side length (LHS) length—i.e. the number of attribute-value pairs in the antecedent (or condition) of a rule; and (3) Overall testing accuracy—i.e. the number of correct classification or predictions made by the rule-set when applied to unseen before objects in the testing set. Rule filtering involves the stepwise elimination of less significant rules from the rule-set based on user-specified filtering criteria, for instance select rules that have RHS support greater than 1. The filtering criteria is progressively increased by fixed steps until the currently reduced set of rules no longer satisfies the user-specified accuracy level. The previous rule-set satisfying the

user-specified accuracy level is finally deemed as the optimum rule-set—i.e. comprising rules with the highest accuracy level, shortest possible length and maximum RHS support.

4. Experimental Results

In this paper we will present experimental results based on two medical datasets: The *Wisconsin Breast Cancer (WBC)* and *New Thyroid Gland (NTG)* datasets. These datasets were chosen for two reasons: (1) all their data vector components being continuous-valued and (2) all the class-subsets are well-separated—i.e. with inter-mean distances fairly large compared to the radii or diameters. The characteristics of the datasets is indicated in Table 1 below.

Table 1. Characteristics of Datasets Used

DATASETS	# OF PATTERNS	# of CLASSES	ATTRIBUTES
New Thyroid Gland	**699** (*683 used* as 16 patterns had missing information)	**3** 1 = Normal 2 = Hyperthyroidism 3 = Hypothyroidism	**5** (*excluding class label*) (1) T3-resin uptake test, (2) Total serum thyroxin, (3) Total serum triiodothyronine, (4) Basal thyroid-stimulating hormone (TSH), (5) Maximal difference of TSH
Wisconsin Breast Cancer	**215**	**2** 0 = Benign 1 = Malignant	**9** (*excluding class label*) (1) Clump thickness, (2) Uniformity of cell size, (3) Uniformity of cell shape, (4) Marginal adhesion, (5) Single epithelial cell size, (6) Bare nuclei., (7) Bland chromatin, (8) Normal nucleoli, (9) Mitoses.

4.1. Phase 1: Data Clustering Using K-Means Clustering Algorithm

Prior to clustering the actual classification information is removed from each dataset—i.e. we work with an un-annotated dataset. The K-means algorithm is used to inductively cluster the data patterns. Upon completion of the clustering process the members of each cluster are associated with their respective class label (see Table 2).

Table 2. Results of K-Means clustering for both datasets

DATASET	CLASS	ACTUAL CLASS DISTRIBUTION (%)	CLUSTER DISTRIBUTION (%)	CLUSTERING ACCURACY (%)
Wisconsin Breast Cancer	Benign	65.01	66.33	**96.1**
	Malignant	34.99	33.67	
New Thyroid Gland	Normal	69.77	70.70	**85.6**
	Hyperthyroidism	16.28	13.02	
	Hypothyroidism	13.95	16.28	

In Hoffman F, Hand DJ, Adams N, Fisher D & Guimaraes G (Eds.) Lecture Notes in Computer Science 2189: Advances in Intelligent Data Analysis, Fourth International Conference (IDA-01),2001, Cascais Portugal. Springer Verlag: Berlin.

4.2. Phase 2: Data Discretization

After the successful clustering of the datasets, we employ the Chi-2 data discretization technique [7] to (1) discretise the continuous data values into meaningful intervals—i.e. nominal values and (2) perform attribute elimination—i.e. attributes that yield only a single discrete value are deemed insignificant and eliminated from the dataset. Table 3 shows discretization results for both datasets. It may be noted that two attributes for each data-set have been eliminated by the data discretisation process. More importantly, we achieved around 65% reduction of data values in both data-sets, consequently yielding a smaller data-set from which cluster-defining symbolic rules are to be generated.

Table 3. Results of data discretisation using the Chi-2 technique. The representation [x, y) means greater than or equal to x but less than y.

WISCONSIN BREAST CANCER DATASET					
Attributes	**Clump Thickness**	**Uniformity Of Cell Size**	**Uniformity Of Cell Shape**	**Marginal Adhesion**	**Single Epithelial Cell Size**
# of Intervals	**2**	**2**	**2**	**0**	**2**
Interval Value	1 = [_ , 7.0) 2 = [7.0, _)	1 = [_ , 3.0) 2 = [3.0, _)	1 = [_ , 3.0) 2 = [3.0, _)	Attribute Eliminated	1 = [_ , 3.0) 2 = [3.0, _)
Attributes	**Bare Nuclei**	**Bland Chromatin**	**Normal Nucleoli**	**Mitoses**	
# of Intervals	**3**	**2**	**2**	**0**	
Interval Value	1 = [_ , 2.0) 2 = [2.0, 8.0) 3 = [8.0, _)	1 = [_ , 4.0) 2 = [4.0, _)	1 = [_ , 3.0) 2 = [3.0, _)	Attribute Eliminated	
NEW THYROID GLAND DATASET					
Attributes	**T3 Resin Uptake**	**Serum Thyroxin**	**Serum Triiodothyronine**	**Basal TSH**	**Absolute Diff. of TSH**
# of Intervals	**5**	**3**	**0**	**2**	**0**
Interval Value	1 = [_, 97.0) 2 = [97.0, 100.0) 3 = [100.0, 118.0) 4 = [118.0, 125.0) 5 = [125.0, _)	1 = [_, 5.7) 2 = [5.7, 14.2) 3 = [14.2, _)	Attribute Eliminated	1 = [_, 4.30) 2 = [4.30, _)	Attribute Eliminated

4.3. Phase 3: Symbolic Rule Discovery

In the next step, symbolic rules were generated from the discretized data-set to explicate the underlying structure of the derived clusters. We randomly partition the dataset into a *training set* (70 % of the dataset) and *testing set* (30 % of the dataset). Since data partitioning is stochastic in nature, we generate *n* randomly partitioned data samples of the same dataset and generate symbolic rule-sets for each sample. Note

that each sample following the same 70:30 distribution of training and testing data, respectively. The eventual rule-set for a dataset will be derived by performing a union of all the rules generated across the *n* number of data samples. For demonstration purposes, we have created 3 partitioned samples of both the datasets.

Symbolic rules were generated from rough set reducts. For pragmatic reasons, the rules discovered were moderated based on two rule-filtering criteria: (1) LHS length and (2) RHS support. Finally, a filtered rule-set was selected based on two criteria: (1) accuracy of the filtered rules when compared with the testing data; and (2) the number of rules in the rule-set—ideally we seek less than 30 rules in the rule-set. In Table 4, we show the rule-sets generated from two separate data samples of the NTG data-set, whereas Table 5 shows the testing results represented as a *prediction matrix*—i.e. the accuracy of the generated rule-sets vis-à-vis the NTG testing data.

Table 4. Exemplar Rule-Sets Generated for the NTG Dataset. The legend implies Cl = Class and Su = RHS Support

	Rule-Set 1 (LHS Length =2)				Rule-Set 2 (LHS Length =2)			
No	**Attributes**		**Cl**	**Su**	**Attributes**		**Cl**	**Su**
1	Basal_TSH(1)	T3_resin(3)	1	87	Basal_TSH(1)	T3_resin(3)	1	89
2	Serum_Thyrox(2)	T3_resin(3)	1	78	Serum_Thyrox(2)	T3_resin(3)	1	83
3	Basal_TSH(1)	T3_resin(1)	2	16	Basal_TSH(1)	T3_resin(1)	2	15
4	Basal_TSH(1)	T3_resin(4)	1	15	Serum_Thyrox(2)	T3_resin(4)	1	13
5	Serum_Thyrox(2)	T3_resin(4)	1	14	Basal_TSH(1)	T3_resin(4)	1	13
6	Serum_Thyrox(3)	T3_resin(1)	2	12	Serum_Thyrox(3)	T3_resin(1)	2	9
7	Basal_TSH(2)	T3_resin(4)	3	8	Serum_Thyrox(1)	T3_resin(4)	3	8
8	Basal_TSH(1)	T3_resin(2)	1	7	Serum_Thyrox(1)	T3_resin(5)	3	8
9	Serum_Thyrox(2)	T3_resin(2)	1	6	Basal_TSH(2)	T3_resin(4)	3	8
10	Serum_Thyrox(3)	T3_resin(3)	1	6	Basal_TSH(2)	T3_resin(5)	3	8
11	Basal_TSH(1)	T3_resin(5)	3	7	Serum_Thyrox(2)	T3_resin(1)	2	6
12	Serum_Thyrox(1)	T3_resin(4)	3	7	Basal_TSH(1)	T3_resin(5)	3	7

Table 5. Prediction Matrix for the Two Rule-Sets Generated From the NTG Data-Set

		0	1	2	Undefined	Accuracy	0	1	2	Undefined	Accuracy
Actual Class	**0**	42	0	0	0	1	41	0	0	3	0.93
	1	2	8	0	0	0.8	0	9	0	2	0.81
	2	1	0	5	7	0.38	2	0	7	1	0.7
	Undefined	0	0	0	0	--	0	0	0	0	--
Average Accuracy						**0.84**					**0.87**

Finally, we get the combined Rule-Set$_{NTG}$ comprising 16 rules, realized via performing the *union* operation over rule-sets 1 and 2—this also illustrates the anticipated overlap between rule-set 1 and 2.

Similar experiments with the WBC data-set yielded 3 separate rule-sets, each derived from 3 randomly distributed samples of the WBC data-set. Each rule-set was then filtered based on LHS length = 5 and RHS support > 20, comprised 29, 26 and

In Hoffman F, Hand DJ, Adams N, Fisher D & Guimaraes G (Eds.) Lecture Notes in Computer Science 2189: Advances in Intelligent Data Analysis, Fourth International Conference (IDA-01),2001, Cascais Portugal. Springer Verlag: Berlin.

11 rules. In Table 6 we present exemplar rules (first 15 only due to lack of space) from rule-set 1 and 2. Again, note the commonalties between rule-sets 1 and 2; the common rules being representative of the dominant class structures inherent within the data. Since the data for class 0 covers almost 65% of the entire data-vector population (see Table 1), the corresponding rules for class 0 have much higher RHS support values, as compared to rules defining class 1. The overall predictive accuracy of the derived rules for both the data-sets, when compared against their respective testing data, was found to be quite high—accuracy for WBC and NTG is about 87% and 86%, respectively—which is a measure of the soundness of our approach.

Table 6. Exemplar Rule-Sets Generated for the WBC Dataset. The Legend is: bn = bare_nuclei, bc = bland_chromatin, ct = clump_thickness, nn = normal_nucleoli, sez = sing_epi_cell_sz, ucp = uni_cell_shape, ucz = uni_cell_size. Cl = Class and Su = RHS Support

	Rule-Set 1 (LHS Length =5)							**Rule-Set 2 (LHS Length =5)**						
No	**Attributes**					**Cl**	**Su**	**Attributes**					**Cl**	**Su**
1	bn(1)	bc(1)	ct(1)	nn(1)	sez(1)	0	245	bn(1)	bc(1)	ct(1)	nn(1)	ucz(1)	0	241
2	bn(1)	bc(1)	ct(1)	nn(1)	ucp(1)	0	243	bn(1)	bc(1)	ct(1)	nn(1)	sez(1)	0	238
3	bn(1)	bc(1)	ct(1)	sez(1)	ucp(1)	0	231	bn(1)	bc(1)	ct(1)	sez(1)	ucz(1)	0	237
4	bn(1)	ct(1)	sez(1)	ucp(1)	ucz(1)	0	231	bn(1)	bc(1)	ct(1)	nn(1)	ucp(1)	0	227
5	bn(1)	ct(1)	nn(1)	sez(1)	ucp(1)	0	229	bn(1)	bc(1)	nn(1)	sez(1)	ucp(1)	0	217
6	bn(1)	bc(1)	nn(1)	sez(1)	ucp(1)	0	228	bn(3)	bc(2)	nn(2)	sez(2)	ucp(2)	1	66
7	bn(3)	bc(2)	nn(2)	sez(2)	ucp(2)	1	62	bn(3)	bc(2)	ct(2)	sez(2)	ucz(2)	1	50
8	bn(3)	ct(2)	sez(2)	ucp(2)	ucz(2)	1	46	bn(3)	bc(2)	ct(2)	nn(2)	ucp(2)	1	43
9	bn(3)	ct(1)	sez(2)	ucp(2)	ucz(2)	1	45	bn(3)	bc(2)	ct(2)	nn(2)	ucz(2)	1	42
10	bn(3)	bc(2)	ct(1)	sez(2)	ucp(2)	1	42	bn(3)	bc(2)	ct(2)	nn(2)	sez(2)	1	39
11	bn(3)	bc(2)	ct(2)	sez(2)	ucp(2)	1	40	bn(3)	bc(2)	ct(1)	sez(2)	ucz(2)	1	34
12	bn(3)	ct(1)	nn(2)	sez(2)	ucp(2)	1	39	bn(2)	bc(2)	nn(2)	sez(2)	ucp(2)	1	30
13	bn(2)	bc(2)	nn(2)	sez(2)	ucp(2)	1	36	bn(3)	bc(2)	ct(1)	nn(2)	ucp(2)	1	29
14	bn(3)	bc(2)	ct(1)	nn(2)	ucp(2)	1	35	bn(3)	bc(2)	ct(1)	nn(2)	ucz(2)	1	28
15	bn(3)	bc(2)	ct(1)	nn(2)	sez(2)	1	35	bn(3)	bc(2)	ct(1)	nn(2)	sez(2)	1	28

5. Concluding Remarks

We conclude that the theory of rough sets proves to be an effective tool for rule discovery because:

- It can extract rules of various granularity, support and coverage
- It does not impose any static statistical parameters or models upon the data, hence minimizing assumptions and allowing the data to represent itself.
- It reduces data by reducing attributes that are both redundant and "unimportant" towards distinguishing between objects and their classes.

We report some interesting observations noted from our experiments:

1. The rough-set approach favors the generation of rules with shorter LHS length. In fact, our experiments demonstrate that as the LHS length decreases the accuracy of the rules increases—i.e. concise rules are more accurate than long rules, which is a desirable effect for knowledge representation purposes.
2. As the RHS support increases both the number of rules and the accuracy tends to reduce. This is in accordance with the theoretical assumptions i.e. as the number of rules reduce less predictive power is available, hence the lower accuracy of the rule-set. This implies the need for a pragmatic balance between an acceptable RHS support value and desired predictive accuracy.

In conclusion we will like to point out that the proposed sequential application of multiple techniques—i.e. data-vector clustering, data discretization, attribute selection and finally rough set approximation—for knowledge extraction via symbolic rule generation, appears to be a pragmatic methodology for the intelligent analysis of un-annotated data-vectors with continuous-valued attributes.

References:

1. Agrawal R., Gehrke J., Gunopulos D., Raghavan P.: Automatic subspace clustering of high dimensional data for data mining applications. Proc. ACM-SIGMOD Int. Conf. Management of Data, Seattle, Washington (1998).
2. Pawlak, Z.: Rough Sets. In: Lin T.Y., Cercone N. (eds.): Rough Sets and Data Mining: Analysis of Imprecise Data. Kluwer Academic Publishers, Dordrecht (1997) pp. 3-7.
3. Abidi, S. S. R., Goh, A., Hoe, K. M.: Specification of Healthcare Expert Systems Using a Multi-Mechanism Rule Extraction Pipeline. Proc. Int. ICSC Congress on Intelligent Systems and Applications, Sydney (2000).
4. Abidi, S.S.R., Hoe, K. M., Goh, A.: Healthcare Simulation Model Specification Featuring Multi-Stage Neural Network Rule Extraction. Proc. 4th Int. Eurosim Congress, Netherlands (2001), to appear.
5. Hu, X., Cercone, N.: Learning Maximal Generalized Decision Rules via Discretization, Generalization and Rough Set Feature Selection. Proc. 9th Int. Conf. on Tools with Artificial Intelligence (1997).
6. Bottou L., Bengio, Y.: Convergence Properties of the K-Means Algorithms. Proc. 7th Int. Conf. on Neural Information Processing Systems, Denver (1994).
7. Liu, H., Setiono, R.: Chi2: Feature Selection and Discretization of Numeric Attributes. Proc. 7th Int. Conf. on Tools with Artificial Intelligence, Washington D.C (1995).
8. Kohavi, R., Sahami, M.: Error-based and Entropy-based Discretization of Continuous Features. Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining (1996) pp. 114-119.
9. Quinlan, R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, California (1993).
10. Tickle A., Andrews R., Golea M., Diederich J.: The Truth Will Come To Light: Directions and Challenges in Extracting the Knowledge Embedded Within Trained Artificial Neural Networks. In: IEEE Trans. on Neural Networks 9:6 (1998) pp1057-1068.
11. Bazan, J. G, Skowron, A. J., Synak, P.: Discovery of Decision Rules from Experimental Data. Proc. 3rd Int. Workshop on Rough Sets and Soft Computing, San Jose CA (1994) pp. 526-533.
12. Bazan J. G.: Dynamic reducts and statistical inference. Proc. 6th Int. Conf. on Information Procesing and Management of Uncertainty in Knowledge-Based Systems (IPMIU'96). Vol. 3, Granada, Spain (1996) pp. 1147-1152.

In Hoffman F, Hand DJ, Adams N, Fisher D & Guimaraes G (Eds.) Lecture Notes in Computer Science 2189: Advances in Intelligent Data Analysis, Fourth International Conference (IDA-01),2001, Cascais Portugal. Springer Verlag: Berlin.

13. Wróblewski, J.: Finding Minimal Reducts using Genetic Algorithms. Proc. 2nd Annual Joint Conf. on Information Sciences, Wrightsville Beach, NC. USA (1995) pp.186-189.
14. Komorowski J., Bjorvand A.T.: Practical Applications of Genetic Algorithms for Efficient Reduct Computation. Proc. of 15th IMACS World Congress on Scientific Computation, Modelling and Applied Mathematics, Berlin (1997).