Available online at www.pelagiaresearchlibrary.com

Pelagia Research Library

Advances in Applied Science Research, 2011, 2 (3): 346-362

ISSN: 0976-8610
CODEN (USA): AASRFC

Diagnosis of Psychopathology using Clustering and Rule Extraction using Rough Set

A Priyadarishini [*], S Karthik, J Anuradha and B K Tripathy

School of Computer Science and Engineering, VIT University, Vellore

---

ABSTRACT

–Psychopathology is a mental stress that tells the level of mental retardation. It is the study of mental illness or mental distress the manifestation of behavior and experience which may indicate the psychological impairment. It is crucial to identify and take necessary action at earlier stage which will improve the quality of life. Psychopathology can be diagnosed using psychometric intelligence scale test i.e. WISC-R and WAIS-R (Wechsler Intelligent Scale Test for Children and Adult) test. An accurate clustering algorithm helps for correct rehabilitation. Soft computing techniques are proposed to apply at different stages of diagnosis. Under data preprocessing Expectation Maximization algorithm is applied for handling missing data. DensityBasedClustering Algorithm gives better result than other clustering algorithm. Outlier analysis is also applied for wrong clusters. For better accuracy in classification, extracting classification rule using Rough Set theory plays a vital role.

Keywords: Density Based Cluster, Missing data analysis, Rough Set, Outlier Analysis, Rule extraction, Wechsler QI scale, Similarity Matrix.

INTRODUCTION

Psychopathology is the study of mental illness. In psychopathological diagnosis a correct classification of a patient's level of mental retardation, especially during childhood and adolescence, is of fundamental importance in order to guarantee appropriate treatment for proper rehabilitation and a quality of life suited to the patient's condition. The methods currently adopted in this field use various diagnostic tools which require time and expertise if they are to be administered correctly. Simplifying the use of these tools would make identification of the most suitable treatment faster and more efficient. For this reason it is needed a technique that is capable to help to execute a faster diagnosis, efficiently and reliably, and at the same time is easy to use for psychologists.

Psychiatrists in particular are interested in psychopathology, which has the aim of describing the symptoms and syndromes of mental illness. The diagnosis of individual patients (to see whether the patient's experience fits any pre-existing classification), or for the creation of diagnostic systems (such as the Diagnostic and Statistical Manual of Mental Disorders or International Statistical Classification of Diseases and Related Health Problems) which define exactly which signs and symptoms should make up a diagnosis, and how experiences and behaviors should be grouped in particular diagnoses. Psychopathology should not be confused with psychopath, which is a type of personality disorder.[1] Psychopathology as a descriptive term may be used to denote behaviors' or experiences which are indicative of mental illness, even if they do not constitute a formal diagnosis. Psychology is the science of mind and behavior. Its immediate goal is to understand behaviors and processes by researching and establishing both general principles and specific cases. Researchers attempt to understand the role of mental functions in individual and social behavior

To meet this need we searched for new methodologies, which we tested on a database comprising data sets of previously diagnosed patients 217 adults. All these patients were administered a psychometric intelligence scale, in a version suitable for their age, i.e. the Italian versions of WAIS-revised [2]. The Wechsler scales are divided subtests that are the basic features of the data set (details about these scales and the database are given in Section 1.1). The data completion problem is often neglected by researchers, especially in psychological field [3], even if it is a problem always present and its wrong handling could be a serious weakness which could lead to inconsistent results. Summarizing, the main aims of the work were the following:

(1) Automatic recognition of the mental retardation level of adults and children by administering Wechsler intelligence scales.
(2) Handles missing data's.
(3) Perform clustering.
(4) Outlier Analysis on wrong clusters.
(5) Rule Extraction and Rule Induction.

1.2. Wechsler intelligence scale

It is one of the most efficient tools for diagnosis cognitive, psychological disorders and it is currently considered as the best tool for measuring intelligence in adults. WAIS-R comprises 11 subtests, 6 included in the Verbal subscale (information, digit span, vocabulary, arithmetic, comprehension and similarities) and five belonging to the performance subscale (picture completion, picture arrangement, block design, object assembly and digit symbol). These subtests are the basic features of the data set [2].

WISC-R is one of the most valuable tools for intelligence assessment in subjects aged 6–16 years, and a valid clinical and diagnostic aid in the area of educational assessment [3].It comprises 12 subtests, six belonging to the verbal subscale (information, similarities, arithmetic, vocabulary, comprehension, digit span) and six to the performance subscale (picture completion, picture arrangement, block design, object assembly, coding, mazes).

The first 10 subtests are called regular tests and the norms of the scale are defined on the basis of these subtests. The remaining subtests (coding and mazes) are called supplementary tests.

Verbal Sub scales

1. Information: 29 questions - a measure of general Knowledge.
2. Digit Span: Subjects are given sets of digits to repeat initially forwards then backwards. This is a test of immediate auditory recall and freedom from distraction.
3. Vocabulary: Define 35 words. A measure of expressive word knowledge. It correlates very highly with Full Scale IQ
4. Arithmetic: 14 mental arithmetic brief story type problems. Tests distractibility as well as numerical reasoning.
5. Comprehension: 16 questions which focus on issues of social awareness.
6. Similarities: A measure of concept formation. Subjects are asked to say how two seemingly dissimilar items might in fact be similar.
7. *Performance WAIS scales*
8. Picture Completion: 20 small pictures that all have one vital detail missing. A test of attention to fine detail.
9. Picture Arrangement: 10 sets of small pictures, where the subject is required to arrange them into a logical sequence.
10. Block Design: Involves putting sets of blocks together to match patterns on cards.
11. Digit Symbol: Involves copying a coding pattern.
12. Object Assembly: Four small jig-saw type puzzles.

II. Literature Survey

Diagnosis of psychological disorder is a challenging task. Here we have tried using clustering algorithms and Rule extraction using Rough Set is discussed in this paper.

1.1 Missing Data

The creation of a tool for a quick and correct classification of mental retardation level, which is needed to choose the right treatment for rehabilitation and to assure a quality of life that is suitable for the specific patient condition .In order to meet this need an adaptive data mining technique that allows to build interpretable models for automatic and reliable diagnosis is essential. Genetic fuzzy system (GFS) is applied to handle missing data. A GFS integrates a classical GA and the fuzzy system which then extended by improving fuzzy to fuzzy C-means (FCM) algorithm and hence called genetic fuzzy C-means (GFCM). The GFCM is cable of not only handling missing data but also able to select the best subset of features to generate an efficient classifier for diagnostic purposes from a database of examples [1]. An accuracy of 79% is obtained when applied to psychological applications.

A simple and effective method for dealing with missing data in decision trees used for classification is MIA. ‘‘Missingness incorporated in attributes” (MIA) is very closely related to the technique of treating ‘‘missing” as a category in its own right, generalizing it for use with continuous as well as categorical variables. A substantial data-based study of classification accuracy shows that MIA exhibits consistently good performance across a broad range of data types and of sources and amounts of missingness. It is competitive with the best of the rest (particularly, a multiple imputation EM algorithm method; EMMI) while being conceptually and computationally simpler [4]. A simple combination of MIA and EMMI is slower but even more accurate.

Mild Cognitive Impairment (MCI) is thought to be the prodromal phase to Alzheimer’s disease (AD), which is the most common form of dementia and leads to irreversible neuro generative damage of the brain [5].In order to further improve the diagnostic quality of the MCI, they developed a MCI expert system to address MCI’s prediction and inference question, consequently, assist the diagnosis of doctor. In this system, mainly deal with following problems: (1) Estimate missing data in the experiment by utilizing mutual information and Newton interpolation (using EM algorithm). (2) make certain the prior feature ordering in constructing Bayesian network. (3) Construct the Bayesian network. The experimental results indicate that Naves Bayes algorithm achieved better results than some existing methods in most instances. The mean square error comes to 0.0173 in the MCI experiment.

2.1.2 Clustering

A genetic algorithm-based clustering technique, called GA-clustering is applied for pattern recognition. The searching capability of genetic algorithms is exploited in order to search for appropriate cluster centers in the feature space such that a similarity metric of the resulting clusters is optimized. The chromosomes, which are represented as strings of real numbers, encode the centers of a xed number of clusters. The *K*-means algorithm, one of the most widely used ones, attempts to solve the clustering problem by optimizing a given metric [6]. The superiority of the GA-clustering algorithm over the commonly used *K*-means algorithm is extensively demonstrated for four artificial and three real-life data sets.

Use of traditional k-mean type algorithm is limited to numeric data. A clustering algorithm based on k-mean paradigm works well for data with mixed numeric and categorical features. A new cost function and Distance measure based on co-occurrence of values are computed. These measures support to identify the significance of an attribute towards the clustering process. A modified description of cluster center is calculated to overcome the limitation of k-mean algorithm (only applicable to numeric data) and provide a better characterization of clusters. The performance of this algorithm has been studied on real world data sets. Comparisons with other clustering algorithms illustrate the effectiveness of this approach. This representation can capture cluster characteristics very effectively, since it contains the distribution of all categorical values in a cluster [7]. The results obtained through this algorithm over a number of real-world data sets are highly encouraging. More sophisticated methods for discretizing numeric valued attributes can definitely better results.

Most partitioning methods cluster objects based on the distance between objects. Such methods can find only spherical-shaped clusters and encounter difficulty at discovering clusters of arbitrary shapes. Other clustering methods have been developed based on the notion of *density*.

Their general idea is to continue growing the given cluster as long as the density (number of objects or data points) in the "neighborhood" exceeds some threshold; that is, for each data point within a given cluster, the neighborhood of a given radius has to contain at least a minimum number of points. Such a method can be used to filter out noise (outliers) and discover clusters of arbitrary shape.

The main goal of density estimation is to find the dense regions of points, which is essentially the same as clustering. Kernel density estimation is a non-parametric technique that does not assume any fixed probability model of the clusters, as in the case of K-means or model-based clustering via the EM algorithm. Instead, kernel density estimation tries to infer the underlying probability density at each point in the dataset. Kernel density estimation over all the *n* points in the datasets essential has $O(n2)$ cost

It can also be shown that K-means is a special case of density-based clustering for an appropriate value of *h* and e. The density attractors correspond to the cluster centroids.Further, it is worth noting that the density-based approach can produce hierarchical clusters,by varying the ẹ threshold. For example, decreasing ẹ can result in the merging of several clusters found at higher thresholds values. In addition, new clusters may emerge if the peak density now satisfies the lower ẹ value.

2.1.3 Rule Extraction

In recent years, support vector machines (SVMs) were successfully applied to a wide range of applications. The classifier is described as a complex mathematical function, it is rather incomprehensible for humans. This opacity property prevents them from being used in many real-life applications where both accuracy and comprehensibility are required, such as medical diagnosis and credit risk evaluation. To overcome this limitation, rules can be extracted from the trained classifier (SVM) that are interpretable by humans and keep as much of the accuracy of the SVM as possible. It provide an overview of the recently proposed rule extraction techniques for SVMs and introduce two others taken from the artificial neural networks domain, being Trepan and G-REX. The described techniques are compared using publicly available datasets, such as Ripley's synthetic dataset and the multi-class iris dataset. Experiments show that the SVM rule extraction techniques lose only a small percentage in Performances compared to SVMs and therefore rank at the top of comprehensible classification techniques [8].

To induce rules from numerical data by rough sets, there are two kinds of methods. One is to discretize the original data and then apply the crisp rough sets models. Here the rough sets models which can only deal with the nominal data are called crisp rough sets models.[9].The other is to fuzzify the original data and then apply fuzzy rough sets models. There are some problems on both of these methods on rules induction such as information loss after discretization or increasing of data size after fuzzification. In this paper we make an attempt to propose one method to induce rules without discretization or fuzzification. Firstly the indiscernibility relation which is the underlining concept of rough sets is redefined as the similarity relation. Subsequently, the concepts of knowledge reduction are proposed based on the similarity relation. Finally, the numerical experiments show that our method is feasible and effective.

III .Architecture and design

This section deals with deals with the proposed architecture and design. In first step of preprocessing, the missing data are handled using EM algorithm, the data are then applied to various classification algorithm.

Fig 3.1describes process of diagnosis mental retardation level using Wechsler intelligence scale. Automatic recognition of the mental retardation level of adults by administering Wechsler intelligence scales. Data set contains 217 adults' records which contains missing value in 31 records. Preprocessing can be done using EM algorithm to obtain complete data.

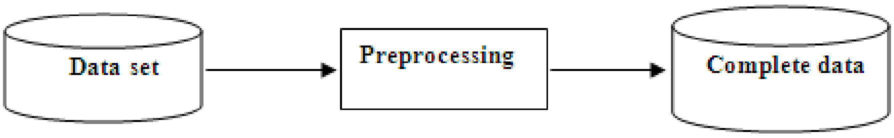


Fig 3.1: preprocessing phase

Fig 3.2 describes clustering phase. On complete dataset MakeDensityBased clustering algorithm is applied to get correct and wrong clusters. Fig 3.3 describes Outlier analysis (K-Means clustering algorithm) on wrong clusters so that it is correctly clustered according to retardation level of patients.

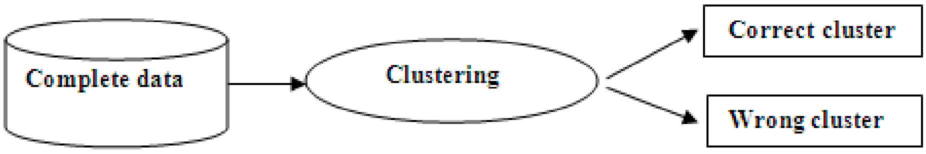


Fig3.2: Clustering phase

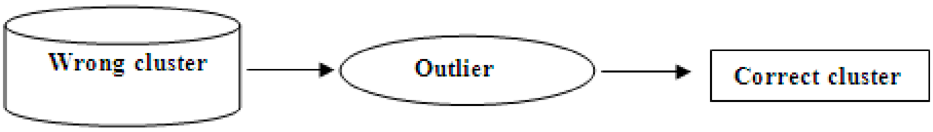


Fig 3.3: Outlier Analysis

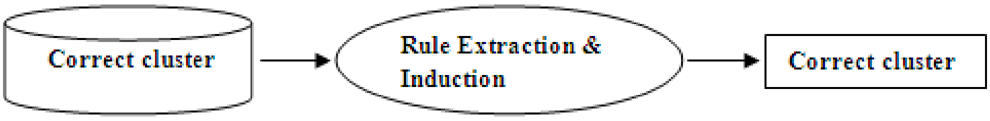


Fig 3.4: Rule Extraction &Induction using Rough Set

Fig 3.4 describes rule extraction and rule induction using Rough Set.50 tuples/objects is taken as input which contain 11 condition attributes and 1 decision attribute.Indiscernability matrix is calculated for condition attributes and Inconsistency vector is calculated and rules have been generated for each tuples. Rule induction is applied on set of rules to get induced rules with all attributes and with attribute reductions. Totally 5 rules have been generated.

IV. Implementation

This section discusses the implementation of the problem

4.1 Preprocessing Phase

Data completion is often a pre-processing statistical technique, in fact there are various statistical methodologies for data completion in the literature (see [8] for a general discussion). The most widely used are two:

1. Regression substitution (RS). This method uses multiple linear regressions to obtain estimates of the missing values. It is applied by estimating a regression equation for each variable, using the others as predictors. Present data is then exploited to obtain the missing data.

2. EM estimation (EME). This is based on the EM algorithm, which comprises two phases: ‘‘E’’ (expectation), predicts an initial value for missing data using other methods (e.g. multiple linear regression); subsequently, in the ‘‘M’’ (maximization) phase the missing values are calculated iteratively using the ‘‘maximum likelihood function’’ until the desired accuracy is reached.

4.1.1 EM algorithm

An expectation-maximization (EM) algorithm is a method for finding maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables. EM is an iterative method which alternates between performing an expectation (E) step, which computes the expectation of the log-likelihood, evaluated using the current estimate for the latent variables, and maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the *E* step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step

Given a statistical model consisting of a set X of observed data, a set of unobserved latent data or missing values Z,and a vector of unknown parameters θ along with a likelihood function $L(\boldsymbol{\theta}; \mathbf{X}, \mathbf{Z}) = p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$, the maximum likelihood estimate (MLE) of the unknown parameters is determined by the marginal likelihood of the observed data

$$L(\boldsymbol{\theta}; \mathbf{X}) = p(\mathbf{X}|\boldsymbol{\theta}) = \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) \qquad \text{-------------} \ (4.1)$$

The EM algorithm seeks to find the MLE of the marginal likelihood by iteratively applying the following two steps:

Expectation step (E-step): Calculate the expected value of the log likelihood function, with respect to the conditional distribution of Z given X under the current estimate of the parameters$\boldsymbol{\theta}^{(t)}$:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathrm{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(t)}}[\log L(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})] \quad \text{------------(4.2)}$$

Maximization step (M-step): Find the parameter that maximizes this quantity:

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \quad \text{-------------(4.3)}$$

First, initialize the parameters $\boldsymbol{\theta}$to some random values.

1. Compute the best value for Z given these parameter values.
2. Then, use the just-computed values of \Z to compute a better estimate for the parameters$\boldsymbol{\theta}$. Parameters associated with a particular value of Z will use only those data points whose associated latent variable has that value.
3. Finally, iterate until convergence.

4.2 Clustering Phase

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. A cluster of data objects can be treated collectively as one group and so may be considered as a form of data compression. Although classification is an effective means for distinguishing groups or classes of objects, it requires the often costly collection and labeling of a large set of training tuples or patterns, which the classifier uses to model each group. It is often more desirable to proceed in the reverse direction: First partition the set of data into groups based on data similarity (e.g., using clustering), and then assign labels to the relatively small number of groups. Additional advantages of such a clustering-based process are that it is adaptable to changes and helps single out useful features that distinguish different groups. Clustering is also called data segmentation in some applications because clustering partitions large data sets into groups according to their similarity. Clustering can also be used for outlier detections.

4.2.1 Make density based Cluster Algorithm

Let $x1$, $x2$, $\cdots$ $,xn$ be a sample of points generated by an underlying probability density function $P(x)$, which is assumed to be unknown. The cumulative distribution function is denoted as $P(x)$.

$$P(x)=\sum_{xi\leq x}p(x_i)$$

Let ˆ $p(x)$ denote the estimate of $p(x)$ at x. We can estimate ˆ $p(x)$ by considering a window of width h centered at x.The width h is a parameter which denotes the spread or smoothness of the density estimate. If the spread is too large we get a more averaged value. If it is too small we do not have enough points in the window. For points within the window ($|z| \leq 1/2$) there is a net contribution of $1/hn$ to the probability estimate $\hat{p}(x)$. On the other hand, points outside the window ($|z| > 1/2$) contribute 0.

$$K(z) = 1/\sqrt{2\pi} \exp\{-z^2/2\}$$

Where $z=x-x_i/2$

Here x (the center of the window) acts as the mean of the distribution, and h acts as the standard deviation of the distribution.

Using this algorithm we obtained 93.86% clustering accuracy. Wrong clusters are grouped for outlier analysis.

4.2 Outlier Analysis

Inconsistent with the set of data, are called outliers. Outlier mining can be described as follows: Given a set of n data points or objects and k, the expected number of outliers, find the top k objects that are considerably dissimilar, exceptional, or inconsistent with respect to the remaining data. Outliers can be caused by measurement or execution error. The outlier mining problem can be viewed as two sub problems:

(1) Define what data can be considered as inconsistent in a given data set, and
(2) Find an efficient method to mine the outliers so defined. Outliers can be caused by measurement or execution error

K-Means clustering algorithm is applied on wrong cluster so that they can fall on correct clusters. An accuracy of 83% is obtained for wrong clusters.

4.3 Rule Extraction and Rule Induction

Let U be the set of objects, which is called by the universe of discourse. A be the set of attributes which is composed of attributes $\{A1, A2, ..., Am\}$. The pair (U, A) is called information system, denoted by IS . If in the information system, we distinguish two classes of attributes, called condition attributes A and decision attributes $A_{m+,1}$. Then such an information system is called decision system, denoted by $DS = AU\,A_{m+1}$

A binary relation $R(B)$ of subset of attributes B satisfying the following two formulas is called *indiscernibility relation:*

R is reflexive, i.e. $R(x, x) = 1$
R is symmetric, i.e. $R(x, y) = R(y, x)$
Let $DS = AU\,A_{m\,1}$ be the decision system, B be the subset of condition attributes,

$INC_{ij}(B)^{A_{m+1}} = R(B)_{ij} - R(A_{m+1})_{ij}$, if $R(B)_{ij} - R(A_{m+1})_{ij} > 0$
----------------------------(4.3.1)

INC_{ij} is called inconsistence degree of object x_i and x. The matrix $INC(B)^{A_{m+1}} = [INC_{ij}(B)^{A_{m+1}}]$ is called inconsistence matrix.

Let $DS = AU\, A_{m+1}$ be the decision system, B be the subset of condition attributes, i.e. $B \subseteq A$ .For a certain threshold, if the following two formulas always hold

$\text{Max}(INC(B)^{A_{m+1}} - INC(A)^{A_{m+1}}) \leq 1-\beta$
----------------------------(4.3.2)

For any attribute b in B , the formula

$\text{Max}(INC(B-\{b\})^{A_{m+1}} - INC(A)^{A_{m+1}}) > 1-\beta$ always holds.

Then the subset of condition attributes B is called β-attribute reduct, denoted by $\text{reduct}(A)^{A_{m+1}}$ Here max(A) denotes the maximum element in matrix A.

Significance of attributes is another important concept of rough sets theory, which is used to measure the discernibility capability of each condition attribute. Here significance of attributes is redefined to measure the discernibility capability of the numerical attributes. This concept is helpful to set up a heuristic algorithm to find the suboptimal attributes reduct.

Let $DS = AU\, A_{m+1}$ be the decision system, B be the subset of condition attributes, $B \subseteq A$. we have

$$SIG(B) = sum(INC(B)^{A_{m+1}})$$

Then $SIG(B)$ is called *significance degree* of subset of attributes *B.*

Each initial rule in the decision table and each object in decision system are symmetric. To reduce each initial rule is equivalent to remove the irrelevant and superfluous attribute-values in each object. Subsequently, the concepts of the reduction of rules are given. For each object x_i in the decision system, it can be represented by an initial rule $attr|^{A}_{i} \rightarrow D_i$.

The former is the conjunction of all the attribute values in condition attributes set A . The latter is the decision attribute-value of the corresponding object.

HEURISTIC ALGORITHM TO FIND SUB OPTIMAL RULE
ALGORITHM-1: To find sub optimal rule of each initial rule.

Input: U --the entire objects set,
x --one object of U ,
A --the entire condition attributes set,

$A_{m,1}$--the decision attribute,
INC --the inconsistence matrix of the entire condition attributes set A with respect to $A_{m,1}$.
Output: Re d the former of the selected reduct rule of x .

Step 1: Initialize Red =ϕ,
Step 2: For each attribute a∪A
T Re d =Re d ∪{a(x)} (Here the notion a(x) represent the value of object x on the attribute a),
Compute the significance degree of attribute-values subset T Re d ,
Select the most significant one
T Re d′=Re d ∪{q(x)},
Re d =T Re d′, A =A∪{q}.
Step 3: For current Re d ,
Compute the inconsistence vector:
INC(Re d) $_{Am+1}$ (X)
Choose the maximum element Max _ e in the vector INC(Re d) $^{Am+1}{}_{i}$- INC $_{i}$.
Step 4: While (A is not empty)
If Max _ e ≤1, β return Re d and stop,
Otherwise go back to step 2 and then step3.

ALGORITHM 2: To induce the rule set

Input: U --the entire objects set,
ERS --the entire suboptimal reduct rules set.
Output: RS is the selected suboptimal rules set.
Step 1: Initialize RS =ϕ
Step 2: While (U is not empty)
For each rule Ru in ERS
Compute the cover number CN and the cover objects set CO of Ru in U ,
Add the rule Ru* which has the biggest cover number CN* to RS ,
ERS =ERS*{Ru},
U =U CO*(Here CO* is the corresponding cover objects set of the rule Ru*).
Step 3: Output RS .

RESULT

In psychological assessment the Wechsler scales are only one of many tools used for diagnosis. Dataset contain 217 records with 17 fields. Dataset containing missing values can be handled by applying EM algorithm.

The following table 5.1 shows the attributes and its type.

Table 5.1 Wechsler intelligence scale attributes

ATTRIBUTES	VALUES
Sex	Nominal(male, female)
Age	Numeric(age in month)
Information	Numeric
Comprehensive	Numeric
Arithmetic	Numeric
Similarities	Numeric
Digit Span	Numeric
Vocabulary	Numeric
Assembly	Numeric
Picture complete	Numeric
picture Arrange	Numeric
Block Design	Numeric
Object Assembly	Numeric
QIVerbal	Numeric
QIPerformance	Numeric
QITotal	Numeric
Retardation level	Nominal(border,moderate,severe)

QIVerbal is sum of all verbal sub tests, QIPerformance is sum of all performance sub tests and QITotal is average of QIVerbal and QIPerformance. Attributes like sex, age, QIVerbal, QIPerformance and QITotal are ignored as input for clustering algorithm.

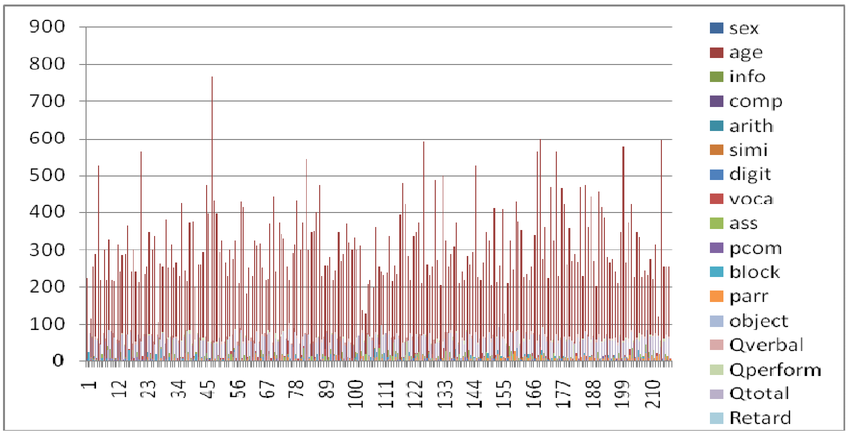
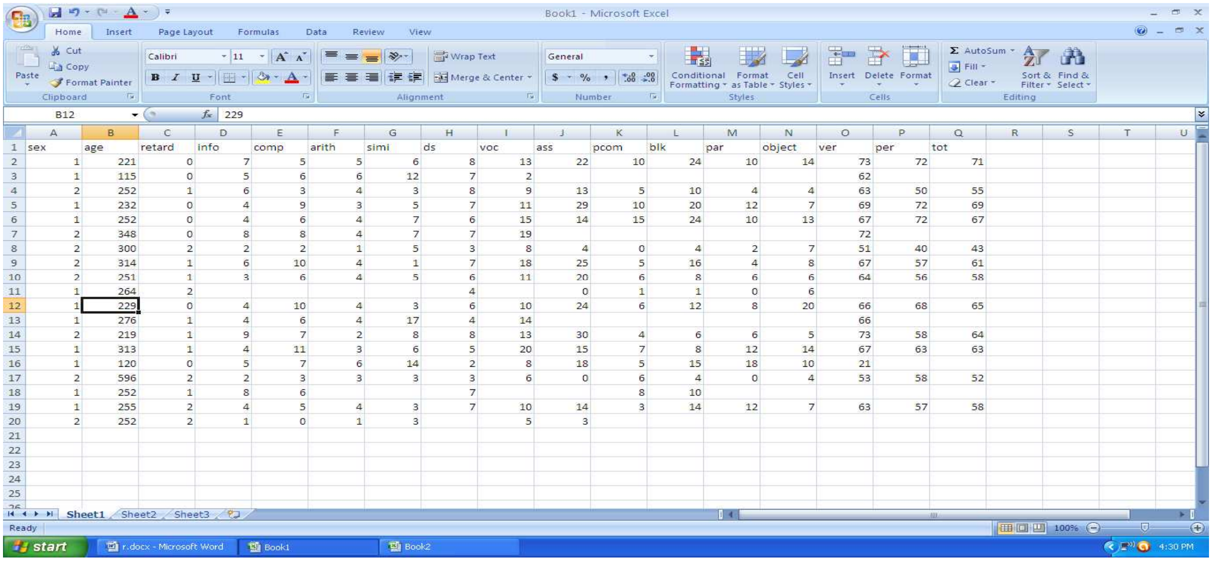


Fig 5.1 Attributes of Wechsler Intelligent Scale

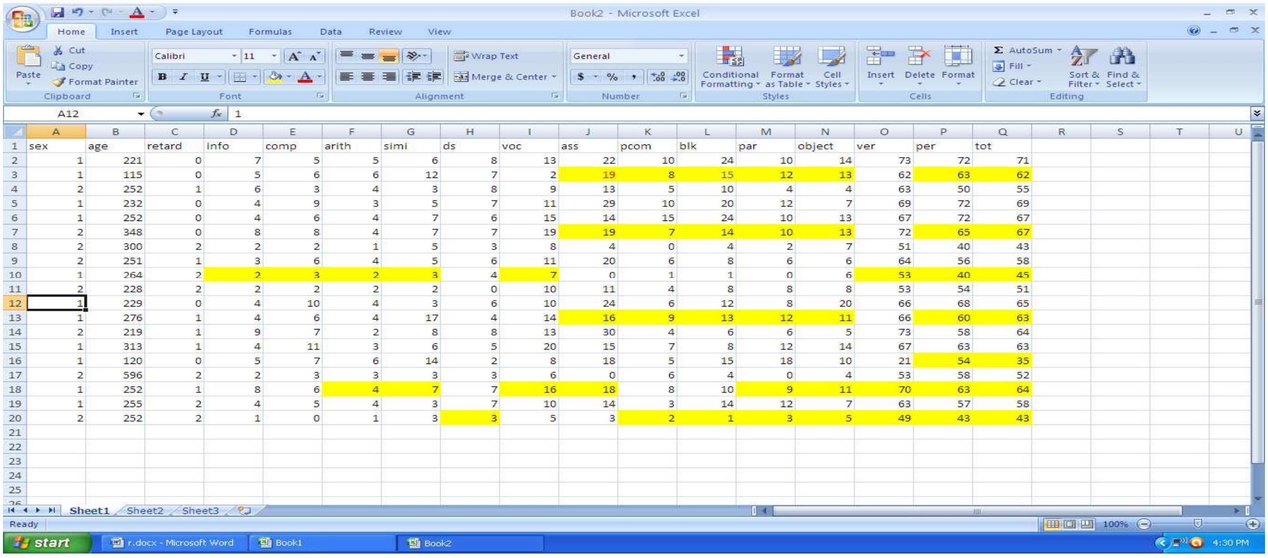
5.1 Preprocessing Phase

Fig 5.1 contains snapshot of missing data's and Fig 5.3 contains snapshot of complete data.217 tuples contains 17 attributes.more than 70 datas are missing.These missing datas are handled using EM algorithm to obtain complete data.SPSS statistical tool is used.



sex	age	retard	info	comp	arith	simi	ds	voc	ass	pcom	blk	par	object	ver	per	tot
1	221	0	7	5	5	6	8	13	22	10	24	10	14	73	72	71
1	115	0	5	6	6	12	7	2						62		
2	252	1	6	3	4	3	8	9	13	5	10	4	4	63	50	55
1	232	0	4	9	3	5	7	11	29	10	20	12	7	69	72	69
1	252	0	4	6	4	7	6	15	14	15	24	10	13	67	72	67
2	348	0	8	8	4	7	7	19						72		
2	300	2	2	2	1	5	3	8	4	0	4	2	7	51	40	43
2	314	1	6	10	4	1	7	18	25	5	16	4	8	67	57	61
2	251	1	3	6	4	5	6	11	20	6	8	6	6	64	56	58
1	264	2					4		0	1	1	0	6			
1	229	0	4	10	4	3	6	10	24	6	12	8	20	66	68	65
1	276	1	4	6	4	17	4	14						66		
2	219	1	9	7	2	8	8	13	30	4	6	6	5	73	58	64
1	313	1	4	11	3	6	5	20	15	7	8	12	14	67	63	63
1	120	0	5	7	6	14	2	8	18	5	15	18	10	21		
2	596	2	2	3	3	3	3	6	0	6	4	0	4	53	58	52
1	252	1	8	6			7			8	10					
1	255	2	4	5	4	3	7	10	14	3	14	12	7	63	57	58
2	252	2	1	0	1	3		5	3							

Fig 5.1.1 Snapshot of missing records



sex	age	retard	info	comp	arith	simi	ds	voc	ass	pcom	blk	par	object	ver	per	tot
1	221	0	7	5	5	6	8	13	22	10	24	10	14	73	72	71
1	115	0	5	6	6	12	7	2	19	8	15	12	13	62	63	62
2	252	1	6	3	4	3	8	9	13	5	10	4	4	63	50	55
1	232	0	4	9	3	5	7	11	29	10	20	12	7	69	72	69
1	252	0	4	6	4	7	6	15	14	15	24	10	13	67	72	67
2	348	0	8	8	4	7	7	19	19	7	14	10	13	72	65	67
2	300	2	2	2	1	5	3	8	4	0	4	2	7	51	40	43
2	251	1	3	6	4	5	6	11	20	6	8	6	6	64	56	58
1	264	2	2	3	2	3	4	7	0	1	1	0	6	53	40	45
2	228	2	2	2	2	2	0	10	11	4	8	8	8	53	54	51
1	229	0	4	10	4	3	6	10	24	6	12	8	20	66	68	65
1	276	1	4	6	4	17	4	14	16	9	13	12	11	66	60	63
2	219	1	9	7	2	8	8	13	30	4	6	6	5	73	58	64
1	313	1	4	11	3	6	5	20	15	7	8	12	14	67	63	63
1	120	0	5	7	6	14	2	8	18	5	15	18	10	21	54	35
2	596	2	2	3	3	3	3	6	0	6	4	0	4	53	58	52
1	252	1	8	6	4	7	7	16	18	8	10	9	11	70	63	64
1	255	2	4	5	4	3	7	10	14	3	14	12	7	63	57	58
2	252	2	1	0	1	3	3	5	3	2	1	3	5	49	43	43

Fig 5.1.2 snapshot of preprocessed data

5.2 Clustering Phase

Upon 217 data's 75% of data's(163 objects) are taken as training set and remaining 25% of data's(54 objects) are taken for testing. MakeDensityBased Cluster algorithm is applied and got three cluster named as 0 (Border), 2 (severe) and 1 (moderate).

Total data: 217
Training set: 163

Test set: 54
No of attributes: 17
Ignored attribute: 5

Training Set
No of cluster: 3
Wrong clustered data: 14
Correct clustered data: 163-14=149
Accuracy=correct clustered data/total data

=149/163 =91%

Table 5.2.1:Clustering Result on Training set datas.

CLUSTER 0		CLUSTER 1	CLUSTER 2
Border level		**Severe level**	**Moderate level**
Actual:	42	67	54
Computed:	43	68	52
Wrong:	3	6	5

Fig 5.2.1 Accuracy of each clusters

TEST SET

No of cluster: 3
Wrong clustered data: 9
Correct clustered data: 54-9=45
Accuracy=correct clustered data/total data
=45/54 =83%

Table 5.2.2: Cluster result for test data's

CLUSTER 0		**CLUSTER 1**	**CLUSTER 2**
	Border level	**Severe level**	**Moderate level**
Actual:	8	30	16
Computed:	6	34	14
Wrong:	3	2	4

Fig 5.2.2: Accuracy of each clusters in testing

Table 5.2.1 and table 5.2.2 shows the cluster result for training as well as testing data's.

5.3 OUTLIER ANALYSIS

Outlier analysis is performed on wrong clusters so that they can fall on correct clusters.K-Means clustering algorithm is used for analysis.
Wrong clusters data: 14+9=23
No of clusters: 3

Table 5.3.1: Accuracy of Outlier

	Cluster 0	**Cluster 1**	**Cluster 2**
	Border	**Moderate**	**Severe**
Actual	7	8	8
Computed	6	8	9

Accuracy: (correct cluster+ correct outlier)/total data

=(149+45+22)/217=215/217 =99%

5.4 RULE EXTRACTION AND RULE INDUCTION

Rule Extraction and Rule induction using Rough set is applied on 30% of datas.The following rules are generated by using this algorithm is:

RULE1:[Info<=1.0]&[comp<=0.0]&[arith<=0.0]&[similar<=1.0]&[ds<=0.0]&[vocabulary<= 5.0]&[assembly<= 0.0]&[pcomp<= 0.0]&[block<= 0.0]&[parrange<= 2.0]&[object<= 2.0]->[retard- 2.0]

RULE2:[Info>=4.0]&[comp>=9.0]&[arith>=3.0]&[similar>=5.0]&[ds>=7.0]&[vocabulary>=11.0]&[assembly>= 29.0]&[pcomp>= 10.0]&[block>= 20.0]&[parrange>= 12.0]&[object>= 7.0]->[retard- 0.0]

RULE3:**[**Info<4.0&info>1.0]&[comp<9.0&comp>0.0]&[arith<3.0&arith>0.0]&[similar<5&similar>1]&[ds<7.0&ds>0.0]&[vocabulary<11.0&vocabulary>5.0]&[assembly<29.0&assembly>0]&[pcomp<10.0&pcomp>0.0]&[block<20.0&block>0]&[parrange<12.0&parrange>2.0]&[object<7.0&object>2.0]->[retard-1.0]

Sub Rules with attribute reduction:

SUBRULE1:[Info>=6.0]&[comp>=9.0]&[arith>=6.0]&[similar>=10.0]&[ds>=10.0]&[vocabulary>=23.0]&[assembly>=39.0]&[pcomp>= 7.0]->[retard- 0.0]

SUBRULE2:[Info<=2.0]&[comp<=3.0]&[arith<=2.0]&[similar<=3.0]&[ds<=4.0]&[vocabulary<=7.0]&[assembly<=0.0]&[pcomp<= 1.0]->[retard- 2.0]

CONCLUSION

Psychopathology is crucial to identify and take necessary action at earlier stage which will improve the quality of life. In automatic diagnosis of mental retardation level using Wechsler intelligence scale MakeDensityBased Clusterer found to outperform the other clustering techniques. Missing data is handled to get complete data which improved the accuracy. An accurate clustering helps for patient’s rehabilitation to lead better life.

An accuracy of 92% is obtained by using MakeDensityBasedCluster. Using Rough Set theory 3 rules and 2 sub rules have been generated by rule extraction and rule induction algorithm with accuracy of 99%.

Acknowledgments
We would like to thank Alessendro G.Di Nuovo for providing the database of patients and giving valuable suggestions.

REFERENCES

[1] Alsseandro G. Di Nuovo a, Vincenzo Catania a, Santo Di Nuovo, Serafino Buono,’ *Applied Soft Computing* 8 (**2008**) Page no: 829–837.

[2] D.Wechsler, WAIS-R Wechsler Adult Intelligence Scale Revised—Manual, The Psychological Corporation, USA, **1981**.

[3] B.E.T.H. Twala , M.C. Jones, D.J. Hand,’ *Pattern Recognition Letters* 29 (**2008**) 950–956

[4] Yan Sun, Yiyuan Tang , Shuxue Ding , Shipin Lv , Yifen Cui.' *Expert Systems with Applications* 38 (**2011**) 442–449.

[5] Ujjwal Maulik, Sanghamitra Bandyopadhyay.' *Pattern Recognition* 33 (**2000**) 1455-1465.

[6] Amir Ahmad, Lipika Dey,' *Data & Knowledge Engineering* 63 (**2007**) 503–527.

[7] Su yun zhaq,Wing W.Y.NG,Eric C,Tsang,Daniel S,Yeung,De-Gang Chen,'Rule induction from numerical data based on rough set theory', Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, 13-16 August **2006**.

[8] S.K. Majumder, A. Gupta, S. Gupta, N. Ghosh,P.K. Gupta, *Journal of Photochemistry and Photobiology B: Biology* 85,(**2006**), Page No: 109–117.

[9] Varun Ramachandran, Anuradha.J, Tisha, Tripathy.V.K , *WCECS* **2010,** vol1541-546

[10] J.L. Schafer, J. Graham, *Psychol. Meth*. 7 (**2002**) 147–177.

[11] Hisao Ishibuchi, Tomoharu Nakashima, Tada Hiko Murata, *Information Science* 136 (**2001**) page No:109-133.

[12] Detlef Nauck, Rudolf Kruse, *Fuzzy Sets and Systems* 89 (**1997**) page no: 277- 288.

[13] Shu-Hsien Liao, Chih-Hao Wen,' *Expert Systems with Applications* 32 (**2007**) 1–11

[14] Z.Pawlak,"Rough Sets Theoretical Aspect of Reasoning about Data", Kluwer Academic Publishers.

[15] Jiawei Han and Micheline Kamber,' Data Mining: Concepts and Techniques, Second Edition.