KNOWLEDGE EXTACRACTION FROM UNCERTAIN DATA-A SURVEY

[1]G.V.Suresh, [2]Dr. E.Srinivasa Reddy

[1]Associate Professor, Dept of CSE, LBRCE, Mylavaram, Andhra Pradesh, India

[2]Professor, Department of CSE, ANUCE, ANU, Andhra Pradesh, India

[1]vijaysuresh.g@gmail.com [2]esreddy67@gmail.com

Abstract: Uncertainty is a ubiquitous element in available knowledge about the real world. Many factors contribute to the uncertainty such as data sampling error, obsolete source, network latency and transmission error, measurement precision limitation ect. These kinds of uncertainty have to be handled cautiously, or else the classification results could be unreliable or even erroneous. There are numerous methodologies developed to comprehend and control uncertainty in data. This paper reviews different methods to handle uncertainty (i.e. Bayesian theory, Certainty Factor, Evidence theory, Possibility theory Fuzzy Set theory and Rough Set theory,) and also purpose hybrid approach which will handle uncertainty to a high level.

Keywords: Data Mining, Uncertain Data, Rough Set Theory, Classification.

1. Introduction

The available knowledge tends to be uncertain in many real applications. Such uncertainties might occur naturally as a result of measurement and decision errors unreliable data transmission and data staling and random nature of the physical data generation and collection process. There are many faces for uncertainty i.e. inconsistency, imprecision, ambiguity, incompleteness, vagueness, unpredictability, noise, unreliability ect. It is significant to handle the uncertainty in various data mining applications to achieve acceptable results [1][2]. Classification is one of the main processes in machine learning and data mining applications. Classification is the method to build a model that can describe and predict the class label of data based on the feature vector [2][5]. An intuitive way of handling uncertainty in classification is to represent the uncertain value by its expectation value and treat it as a certain data.We usually make decisions based on incomplete and partially inaccurate data. For example, we purchase goods without learning about all their features, sign contracts without fully understanding the fine print, and make investments without full knowledge of the financial markets[7][8]. Human experts almost never have complete data relevant to their work, and they usually can make reasonable decisions in the face of uncertainty. Many approaches have been developed to handle uncertainty like rough set theory, fuzzy set theory, evidence theory, possibility theory, Bayesian theory, statistical functions, and certainty factor [9]. But all these theories have some advantages and disadvantages. These theories are somehow valid for some specific purpose only i.e. each technique is applicable to a particular problem only. Therefore, it is essential to develop a hybrid approach that is composed of a combination of two or more theories. Shortcomings in individual theories can be mitigated by integrating various methods. This study first attempt to provide a better understanding of uncertainty by giving a comprehensive overview of its classifications and theories to handle uncertainty till now and also describes a hybrid approach to handle uncertainty to a high level

2. Research Background

Uncertainty can be due to lack of knowledge or insufficient information. This can be due to vagueness, no specificity and conflict in the information. Data is said to be perfect when they are precise and certain. There are many faces for uncertainty i.e. inconsistency, imprecision, ambiguity, incompleteness, vagueness, unpredictability, noise, unreliability ect[4][5]. Human experts almost never have complete data relevant to their work due to lack of adequate data, inconsistency of data, inherent human fuzzy concept, matching of similar situations, different opinions, ignorance, lack of available theory to describe a particular situation and they usually can make sensible decisions in countenance of uncertainty [3].

2.1 Sources of Uncertainty

There are several sources of uncertainties i.e. uncertainty can arise from various sources such as: some of them are

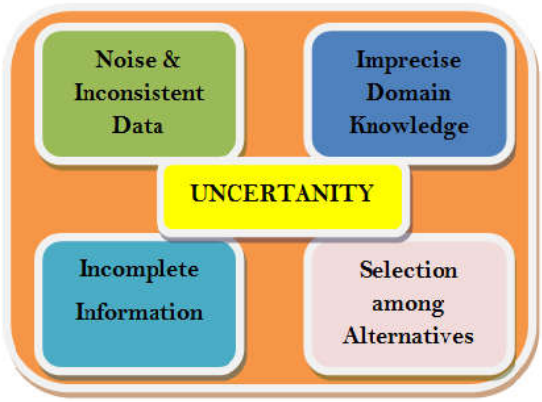

Figure 1. Some sources of uncertanity

2.1.1 Noisy and inconsistent Data

Real world data is unclean. When data is collected from different sources, then there are many chances that the data is noisy and conflicting. Noisy data usually refers to the meaningless data whereas inconsistent data refers to the ambiguous data i.e. If a coin is flipped, then output has to be either head or tail but it cannot be a number(Noisy data) & salary of a person is 5000 and it lies between 1000-4000 (inconsistent Data)

2.1.2 Incomplete Information

Incomplete Information is partial or insufficient information. Information may become incomplete due to the changes in factors with time i.e. classification can be done accurately only if complete information is available.

2.1.3 Imprecise Knowledge Domain

Imprecise Knowledge Domain refers to the knowledge which is not complete in a particular domain. The result of incomplete domain knowledge may or may not always results in. i.e. A doctor must have the complete knowledge about a patient's disease, its history, period of suffering or about symptoms of disease etc. if he has Imprecise Knowledge Domain, than he might not be able to treat him accurately.

2.1.4 Selection among the Alternatives

There are many alternatives to a problem and there may be a disagreement in choosing the best alternative among all. It is always not necessary that the best solutions comes out for a solving a problem. i.e There is various methods to handle uncertainty. So choosing the best way to handle it requires a complete understanding about the problem as well as the methods and then applying the appropriate

2.2 Types of Uncertainty

There are mainly two types of uncertainty as proposed by Regan (2002).The first one is called as Epistemic Uncertainty (indeterminate facts) and other one is Linguistic Uncertainty (indeterminacy in language).

2.2.1 Epistemic Uncertainty

The following can be considered for the causes of epistemic uncertainty

2.2.1.1 Systematic Error

This error occurs due to bias or miscalibration in experiment while doing some experimentation

2.2.1.2 Measurement Error

This error occurs due to ambiguity or inaccuracy in the measurement procedure itself

2.2.1.3 Natural Disparity

This will occurs due to the uneven changes in the environment

2.2.1.4 Inherent Uncertainty

This is due to the factors contributing Systematic error, Measurement error Natural disparity or combination of them

2.2.1.4 Imperfect Data

The data or information used in decision making procedure may be corrupt and imperfect which further leads to wrong decision.

2.2.2 Linguistic Uncertainty

2.2.2.1Vagueness

Vagueness is connected with such concepts as fuzziness, indistinctiveness, haziness, cloudiness, unclearness, and shapelessness [1][2][3]. Vagueness reveals the difficulty of making sharp or precise distinctions in the world i.e. the information available in the model is unclear for example considered the statement: “I will see you at 4:00 p.m. or 5:00 p.m.” is a vague statement.

2.2.2.2 Ambiguity

Ambiguity is connected with such concepts as non-specificity, one-to-many relations, variety, generality, diversity and divergence i.e. some elements of model lack complete semantics, leading to several possible interpretations for example considered the statement: “Kamala is in early old age” is a ambiguous statement[2].

Context dependence

Context dependence is uncertainty arising from a failure to specify the context in which a term is to be understood [6]. A clear way to deal with context dependence is to specify such a context. Under-specificity occurs when there is unwanted generality; the statement in question does not provide the degree of specificity required in order to proceed with an assessment or decision [9].

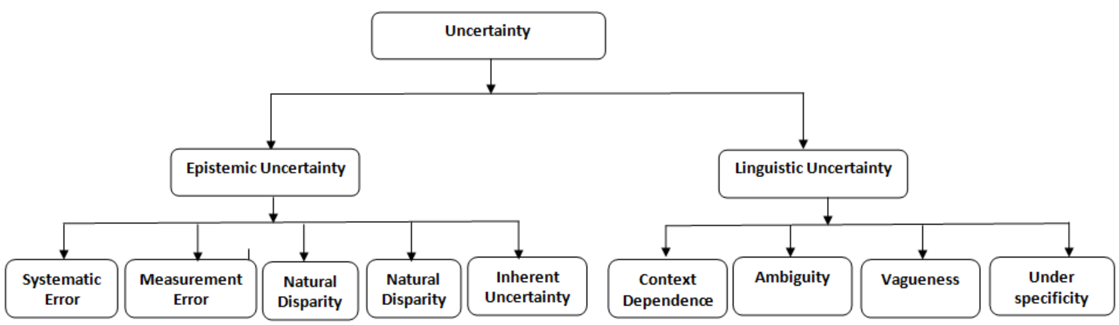

Figure 2. Types of uncertanity

2.3 Some Approaches to Handle Uncertainty

2.3.1 Basic concepts of probability

Probability theory is a branch of mathematics concerned with the analysis of random phenomena. The outcome of a random event cannot be determined before it occurs, but it may be any one of several possible outcomes [10][12]. The actual outcome is considered to be determined by chance. There are many probability theory formulas like conditional probability, conditional independence, Baye’s theorem, etc that can be used for analyzing [13]. Probability is often used to handle uncertainty arising from incomplete information. Probability lies in between 0 and 1. Some of the axioms of probability are [12]:

Let S be a sample space, A and B are events.

- All values of probabilities are between zero and one
- Probabilities of an event that are necessarily true have a value of one i.e. P(True)=1,and those that are necessarily false have a value of zero i.e. P(False)=0.
- The probability of a disjunction is given by
- Calculating the negation of a probability of an event is given by P(A) = 1-P(A)

2.3.2 Joint Probability

Joint Probability of the occurrence of two independent events is written as **P (A and B)** and is defined by

$$P\ (A \text{ and } B) = P\ (A \cap B) = P\ (A) * P\ (B)$$

Example: We toss two fair coins separately.

Let P (A) = 0.5, Probability of getting Head of first coin

P (B) = 0.5, Probability of getting Head of second coin

Probability (Joint probability) of getting Heads on both the coins is

$$P\ (A) * P\ (B) = 0.5*0.5 \quad = 0.25$$

The probability of getting Heads on one or on both of the coins i.e. the union of the probabilities P (A) and P (B) is expressed as

$$P\ (A \text{ or } B) = P\ (A \cup B) = P\ (A) + P\ (B) - P\ (A) * P\ (B)$$
$$= 0.5 + 0.5 - 0.25$$
$$= 0.75$$

2.3.3 Conditional Probability

Suppose a rational agent begins to perceive data from its world, it stores this data as evidence. This evidence is used to calculate a posterior or conditional probability which will be more accurate than the probability of an atomic event without this evidence (known as a prior or unconditional probability)[13][14].Probability of an event H (Hypothesis), given the occurrence of an event E (evidence) is denoted by P(H | E) and is defined as follows:

Example: What is the probability of a person to be male if person chosen at random is 80 years old?

The following probabilities are given

- Any person chosen at random being male is about 0.50
- probability of a given person be 80 years old chosen at random is equal to 0.005
- probability that a given person chosen at random is both male and 80 years old may be =0.002

The probability that an 80 years old person chosen at random is male is calculated as follows:

P(X is male | Age of X is 80)

= [P(X is male and the age of X is 80)] / [P (Age of X is 80)]

= 0.002 / 0.005 = 0.4

2.4 Bayes Theorem for handling Uncertainty

Bayes theorem provides a mathematical model for this type of reasoning where prior beliefs are combined with evidence to get estimates of uncertainty. This approach relies on the concept that one should incorporate the prior probability of an event into the interpretation of a situation. It relates the conditional probabilities of events.

It allows us to express the probability P (H | E) in terms of the probabilities of P (E |H), P (H) and P (E).

Example:

Find whether Bob has a cold (hypotheses) given that he sneezes (the evidence) i.e. calculate P (H | E). Suppose that we know given the following.

P (H) = P (Bob has a cold) = 0.2

P (E |H) = P (Bob was observed sneezing = 0.75

P (E |~H) = P (Bob was observed sneezing = 0.2

P (H | E) = P (Bob has a cold | Bob was observed sneezing) = [P (E | H) * P (H)] / P (E)

P (E) = P (E and H) + P (E and ~H)

= P (E | H) * P (H) + P (E | ~H) * P (~H)

= (0.75) (0.2) + (0.2) (0.8) = 0.31

P (H | E) = [(0.75 * 0.2)] / 0.31 = 0.48387

We can conclude that “Bob’s probability of having a cold given that he sneezes” is about 0.5. Further it can also determine what is his probability of having a cold if he was not sneezing?

P (H |~E) = [P (~E | H) * P (H)] / P (~E)

= [(1 – 0.75) * 0.2] / (1 – 0.31)

=0.05 / 0.69 = 0.072

Hence Bob’s probability of having a cold if he was not sneezing” is 0.072

2.4.1 Problems with Probabilistic Approach

- Requires large amounts of probability data (Sufficient Sample Sizes)
- Inexact or incorrect especially for subjective probabilities and subjective evidence may not be reliable.
- Independence of evidences assumption often not valid
- Relationship between hypothesis and evidence is reduced to a number
- High computational overhead
- Data can't readily be reduced to numbers or frequencies.
- Human estimates of probabilities are notoriously inaccurate

2.5 Certainty Factor for handling Uncertainty

Standard statistical methods are based on the assumption that an uncertainty is the probability that an event (or fact) is true or false. Certainty theory is a framework for representing and working with degrees of belief of true and false in knowledge-based systems [4][13]. In certainty theory uncertainty is represented as a degree of belief. There are two steps in using every non-probabilistic method of uncertainty. First, it is necessary to be able to express the degree of belief. Second, it is necessary to manipulate (e.g., combine) degrees of belief when using knowledge-based systems. Certainty factors are not probabilities, they represent beliefs about how strong given evidence is, and to what degree the evidence (E) supports a hypothesis (H). Certainty factors are measured using various scales (0 – 100, 0 – 10, 0 – 1, -1 to- 1) and linguistics ones (certain, fairly certain, likely, unlikely, highly unlikely, definitely not).Usually three cases are considered [13]

- When CF = +1, indicate strong confidence in a hypothesis (H)
- When CF = 0, it means hypotheses H is not known
- When CF = -1, indicate confidence against a hypothesis (H)

CFs are calculated using two other measures

i) MB (H, E) – Measure of Belief: value between 0 and 1 representing the degree to which belief in the hypothesis H is supported by observing evidence E. Measure of Belief can be calculated using

$$MB(H,E)=\begin{cases}\frac{P(H|E)-P(H)}{1-P(H)} & \text{if } P(H|E)\geq P(H) \text{and } P(H)\neq 1\\ 1 & \text{if } P(H)=1\\ 0 & \text{if } P(H|E)<P(H)\end{cases} \quad (1)$$

ii) MD (H, E) – Measure of Disbelief: value between 0 and 1 representing the degree to which disbelief in the hypothesis H is supported by observing evidence E.

$$MD(H,E)=\begin{cases}\frac{P(H)-P(H|E)}{P(H)} & \text{if } P(H|E)\leq P(H) \text{and } P(H)\neq 0\\ 1 & \text{if } P(H)=0\\ 0 & \text{if } P(H|E)>P(H)\end{cases} \quad (2)$$

CF is calculated in terms of the difference between MB and MD:

$$CF(H,E)=MB(H,E)-MD(H,E) \quad (3)$$

2.5.1 Combining Certainty Factors

Multiple sources of evidence produce CFs for the same fact.

$$MB(P1ANDP2)=MIN\left(MB\left(P1\right),MB\left(P2\right)\right) \quad (4)$$

$$MB(P1ORP2)=MAX\left(MB\left(P1\right),MB\left(P2\right)\right) \quad (5)$$

Negation of a fact can be represented by

$$MB(NOT\ P1)=1-MB(P1) \quad (6)$$

2.5.2 Rules with Uncertain Evidence

To deal with uncertainty in rules, CF uses credibility. The credibility is then multiplied by the MB for the conclusion of the rule.

$$MB(\text{Conclusion}) = MB(\text{Conditions}) * \text{Credibility} \quad (7)$$

$$MB[H:E1,E2] = MB[H:E1] + MB[H:E2]*(1-MB[H:E1]) \quad (8)$$

Consider an example

Rule 1 IF Interest Rate Low
AND Inflation is Low
THEN X Buy Stocks
Rule 2 IF Market Steady
OR Employment Good
THEN X Buy Stocks
Rule 3 IF Inflation High
OR Employment Not Good
THEN X don't buy stocks

Interest Rate Low	0.9
Inflation Low	0.7
Market Steady	0.8
Employment Good	0.6
Inflation High	0.7

While the credibility of each rule is as follows

Rule 1	0.7
Rule 2	0.8
Rule 3	0.6

With this inputs test the hypothesis whether "X Buys the Stocks"
MB[X Buys the Stocks: Rule 1] = MIN (0.9, 0.7) * 0.7 = 0.49
MB[X Buys the Stocks: Rule 2] = MAX (0.8, 0.6) * 0.8 = 0.64
MB[X Buys the Stocks: Rule 3] = MAX (0.7), (1-0.6)) * 0.6 = 0.42
Combining the Rule 1 and Rule 2:
MB[X Buys the Stocks: Rule1, Rule2] = MB[X Buys the Stocks: Rule1] + MB[X Buys the Stocks: Rule 2]*(1-MB[X Buys the Stocks: Rule1]) = 0.49 + 0.64 * (1 - 0.49) = 0.82
Combining the three rules:
CF[X Buys the Stocks: Rule 1, Rule 2, Rule 3]
= MB[X Buys the Stocks: Rule 1, Rule 2] - MD[X Buys the Stocks: Rule 3]
= 0.82 - 0.42
= 0.4

After getting the CF for the hypothesis, what is the answer for the question "X Buys the Stocks"? In an expert system that implements uncertainty handling would answer by considering uncertain terms and their interpretation in MYCIN "May be"[13]. Isn't it exactly the way say it? That's why Certainty Factor has been criticized to be excessively ad-hoc.The semantic of the certainty value can be subjective and relative.

2.5.3 Difficulties with Certainty Factors

- In some cases results may depend on the order in which evidences are considered.
- Combination of non-independent evidences is unsatisfactory
- New knowledge may require changes in the certainty factors of existing knowledge
- Not suitable for long increase chains

2.6 Dempster-Shafer Theory of Evidence for handling Uncertainty

The Dempster-Shafer theory is a mathematical theory of evidence based on belief functions and plausible reasoning, which is used to combine separate pieces of information (evidence) to calculate the probability of an event[15][16]. In the D-S theory of evidence, the set of all hypotheses that describes a situation is the frame of discernment. The hypotheses should be mutually exclusive and exhaustive, meaning that they must cover all the possibilities and that the individual hypotheses cannot overlap. The D-S theory mirrors human reasoning by narrowing its reasoning gradually as more evidence becomes available. Two properties of the D-S theory permit this process:

- the ability to assign belief to ignorance
- the ability to assign belief to subsets of hypotheses

Two special sets applied in D-S theory. The first is the null set, which cannot hold any value and the second special set is a set contains all elements. Assigning belief to the second set does not help distinguish anything and representing ignorance [17][18]. Humans often give weight to the hypothesis "I don't know", which is not possible in classical probability. Assigning belief to "I don't know" allows us to delay a decision until more evidence becomes available. Each data source, Si for example, will contribute its observation by assigning its beliefs. This assignment function is called the "probability mass function" and denoted by mi. So, the upper and lower bounds of a probability interval can be defined as contains the precise probability of a set of interest in the classical sense, and is called belief and plausibility [20]. The lower bound of the confidence interval is the belief confidence, which accounts all evidences E_k that support the given proposition "A":

$$\text{Belief}_i(A) = \sum_{E_k \subseteq A} m_i(E_k)$$

The upper bound of the confidence interval is the plausibility confidence, which accounts all the observations that does not rule out the given proposition:

$$\text{Plausibility}_i(A) = 1 - \sum_{E_k \cap A = \varphi} m_i(E_k)$$

For each possible proposition (e.g., A), Dempster-Shafer theory gives a rule of combining data source S_i's observation mi and data source S_j's observation m_j

$$(m_i \text{ Å } m_j)(A) = \frac{\sum_{E_k \cap E_{k'} = A} m_i(E_k)\, m_j(E_{k'})}{1 - \sum_{E_k \cap E_{k'} = j} m_i(E_k)\, m_j(E_{k'})}$$

The Dempster's rule of combination is a generalization of Bayes' rule [16][17]. This rule strongly emphasizes the agreement between multiple sources and ignores all the conflicting evidence through a normalization factor. Compared with Bayesian theory, the Dempster-Shafer theory of evidence is much more analogous to our human perception-reasoning processes. Its capability to assign uncertainty or ignorance to propositions is a powerful tool for dealing with a large range of problems that otherwise would be intractable.

Let $\Theta = \{E_1, \neg E_1\}$

The measures of uncertainty, taken collectively are known in Dempster-Shafer Theory terminology as a "basic probability assignment" (bpa). Hence we have a bpa, say

$m_1(E_1) = 0.8$ $m_2(E_1) = 0.9$

$m_1(\neg E_1) = 0$ $m_2(\neg E_1) = 0$

$m_1(E_1, \neg E_1) = 1 - m_1(E_1) = 1 - 0.8 = 0.2$

$m_2(E_1, \neg E_1) = 1 - m_2(E_1) = 1 - 0.9 = 0.1$

$m_1 \downarrow$	$m_2 \rightarrow$ E_1=0.9	$\neg E_1$=0	$E_1, \neg E_1$=0.1
E_1=0.8	0.72	0	0.08
$\neg E_1$=0	0	0	0
$E_1, \neg E_1$=0.2	0.18	0	0.02

Calculate K by using the formula

$$K^{-1} = 1 - \sum_{X \cap Y = \varnothing} m_1(X) \bullet m_2(Y)$$

There are only two cases where X Y=Ø. Case (i) $m_1(E_1) \bullet m_2(\neg E_1)$ (ii) $m_1(\neg E_1) \bullet m_2(E_1)$

$$K^{-1} = 1 - m_1(E_1) \bullet m_2(\neg E_1) + m_1(\neg E_1) \bullet m_2(E_1) = 1-(0+0) = 1$$

Combing m1 and m2 can be done using

$$m_i(A) = K \sum_{X \cap Y = A} m_1(X) \bullet m_2(Y)$$

$m_1 \oplus m_2(E_1) = 1 \bullet (0.72+0.08+0.18) = 0.98$

$m_1 \oplus m_2(\neg E_1) = 1 \bullet (0) = 0$

$m_1 \oplus m_2(E_1, \neg E_1) = 1 \bullet (0.02) = 0.02$

From the above evidence we can conclude that most probable belief is $E_1$and evidence given by m_1, m_2 is not contradict.

2.6.1 Dempster-Shafer Pros and Cons

Dempster-Shafer Theory Pros	Dempster-Shafer Theory Cons
Addresses questions about necessity and possibility that Bayesian approach cannot answer	Source of evidence not independent; can lead to misleading and counter-intuitive results
Prior probabilities not required, but uniform distribution cannot be used when priors are unknown	The normalization in Dempster's Rule loses some meta information, and this treatment of conflicting evidence is controversial
Useful for reasoning in rule-based systems	Bayesian approach can also do something akin to confidence interval by examining how much one's belief would change if more evidence acquired

2.7 Fuzzy Set Theory/Logic

Fuzzy sets were introduced by Lotfi Zadeh in 1965 to represent/manipulate data and information possessing of non-statistical uncertainties [20][21]. Fuzzy sets are a precise mathematical tool for processing data that is derived from uncertain and vague sources. We can define uncertainty in view of fuzzy set theory as follows: When A is a fuzzy set and x is a relevant object, the proposition "x is a member of A" is not necessarily either true or false. It may be true only to some degree, the degree to which x is actually a member of A. For example: "the weather outside is cold"

- But, how cold is actually the coldness you described?
- What do you mean by 'cold' here?

As we can see a particular temperature is cold to one person but it is not to another. Fuzzy logic was introduced to design systems that can demonstrate human-like reasoning capability to understand such vague terms. Boolean Logic (for 'Temperature') to describe terms such as 'cold' ,'hot'

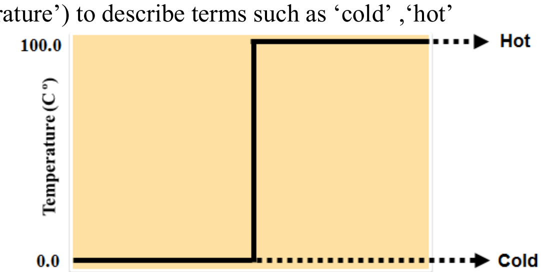

Figure 3.Boolean Logic (for 'Temperature')

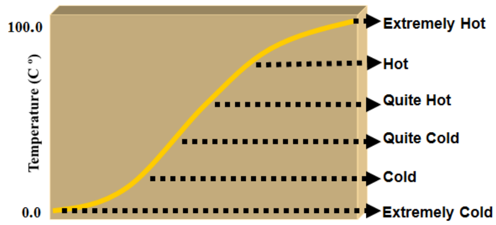

Figure 4. Fuzzy Logic (for 'Temperature')

Unlike two-valued Boolean logic, fuzzy logic is multi-valued. It deals with degrees of membership and degrees of truth. This distinctions can made more exact by the membership function. The membership function of a fuzzy set A is denoted by

$$\mu_A(x): X \rightarrow [0,1]$$ where

$\mu_A(x) = 1$ 1 if x is totally in A;

$\mu_A(x) = 0$ if x is not in A

$0 < \mu_A(x) < 1$ if x is partly in A

This set allows a continuum of possible choices. For any element x of universe X, membership function $\mu_A(x)$ equals the degree to which x is an element of set A. This degree, a value between 0 and 1, represents the degree of membership, also called membership value, of element x in set A [22].Let us consider "tall men" example the universe of discourse – the men's heights – consists of three sets: short, average and tall men. As you will see, a man who is 184 cm tall is a member of the average men set with a degree of membership of 0.1, and at the same time, he is also a member of the tall men set with a degree of 0.4. Typical functions can be used to represent a fuzzy set i.e. sigmoid, Gaussian and pi. However, these functions increase the time of computation. Therefore, in practice, most applications use linear fit functions.

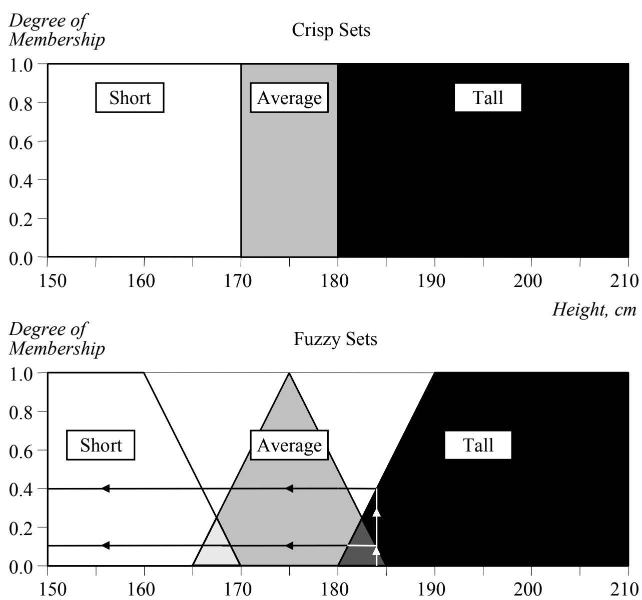

Figure 5. Fuzzy sets representation for "tall men

2.7.1 Linguistic Variables & Hedges

A linguistic variable is a fuzzy variable [22]. For example, the statement "John is tall" implies that the linguistic variable John takes the linguistic value tall. The range of possible values of a linguistic variable represents the universe of discourse of that variable. For example, the universe of discourse of the linguistic variable speed might have the range between 0 and 220 km/h and may include such fuzzy subsets as very slow, slow, medium, fast, and very fast. A linguistic variable carries with it the concept of fuzzy set qualifiers, called hedges. Hedges are terms that modify the shape of fuzzy sets. They include adverbs such as very, somewhat, quite, more or less and slightly. Linguistic variables are also used in fuzzy rules.

For example:

IF wind is strong
THEN sailing is good

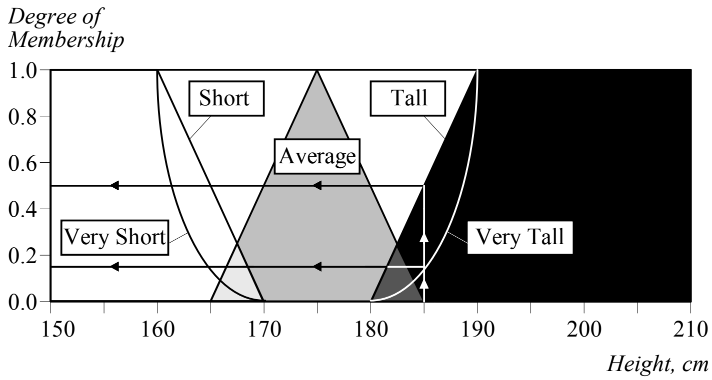

Figure 6. Linguistic variable representation for “tall men”

Linguistic variable representation for “tall men”

Some of the common hedges used in many applications

- Very à $f(x)^2$
- Not very à $1 – f(x)^2$
- Somewhat à $f(x)^{1/2}$
- About (or around) à f(x) +/- delta
- Nearly à f(x) – delta

2.7.2 Fuzzy Logic Process

First, inputs are translated into fuzzy values [23][24]. This process is sometimes referred to as fuzzification. For instance, assume Sam is 21, John is 42 and Frank is 53, we want to know who is “old” and who is “young”

- Young = { Sam / .7, John / .2, Frank / .1}
- Old = { Sam / .1, John / .4, Frank / .6}

Next, we want to infer over our members

We might have rules, for instance that you cannot be a country club member unless you are OLD and WEALTHY

The rules, when applied, will give us conclusions such as Frank can be a member at .8 and Sam at .5

Finally, given our conclusions, we need to defuzzify them. There is no single, accepted method for defuzzification

- How do we convert Frank / .8 and Sam / .5 into actions?
- Often, methods compute “centers of gravity” or a weighted average of some kind,
- The result is then used to determine what conclusions are acceptable (true) and which are not

Example

Consider the rules for Air-conditioner

- if air is warm and dry, decrease the fan and increase the coolant
- if air is warm and not dry, increase the fan
- if air is hot and dry, increase the fan and the increase the coolant slightly
- if air is hot and not dry, increase the fan and coolant
- if air is cold, turn off the fan and decrease the coolant

Our input obviously requires the air temperature and the humidity, the membership function for air temperature. The following rules can be drawn from the above rules

if it is 60, it would be considered cold 0, warm 1, hot 0

if it is 85, it would be cold 0, warm .3 and hot .7

Temperature = 85, humidity is moderately high

hot .7, warm .3, cold 0, dry .6, not dry .4

Rule 1 has “warm and dry”

- warm is .3, dry is .6, so “warm and dry” = min(.3, .6) = .3

Rule 2 has “warm and not dry”

- min(.3, .4) = .3

Rule 3 has "hot and dry" = min(.7, .3) = .3

- our fourth and fifth rules give us 0 since cold is 0

Our conclusions from the first three rules are to

- decrease the coolant and increase the fan at levels of .3
- increase the fan at level of .3
- increase the fan at .3 and increase the coolant slightly

To combine our results, we might increase the fan by .9 and decrease the coolant (assume "increase slightly" means increase by ¼) by .3 - .3/4 = .9/4

Finally, we defuzzify "decrease by .9/4" and "increase by .9" to actionable amounts

2.7.3 Problems with Fuzzy Set Theory

There are no learning mechanisms in fuzzy logic

What if we have membership functions provided from two different people

- for instance, what a 6'11" basketball player defines as tall will differ from a 4'10" gymnast

Membership values begin to move away from expectations when chains of logic are lengthy so this approach is not suitable for many KBS problems (e.g., medical diagnosis)

- The theory is not concerned about how the rules are created, but how they are combined
- The rules are not chained together, instead all fire and the results are combined

2.8 Possibility Theory

Possibility theory is a mathematical theory devoted to the handle uncertainties in data and is an alternative to probability theory [25]. Possibility theory has enabled a typology of fuzzy rules, either in the background knowledge on which possibility is based, or in the set for which possibility is asserted. This constitutes a method of formalizing non-probabilistic uncertainties on events: i.e. a mean of assessing to what extent the occurrence of an event is possible and to what extent we are certain of its occurrence, without knowing the evaluation of the possibility of its occurrence. For example, imprecise information such as "Room temperature is above 28 degrees" implies that any temperature T above 28 is possible and any temperature equal to or below 28 is impossible. This can be represented by a "possibility" measure defined on the temperature domain whose value is 0 (impossible) if T<28 and 1 (possible) if T≥28. When the predicate is vague like in "Room is cool", the possibility can be accommodate degrees, the largest the degree, the largest the possibility. There are two approaches to possibility theory: one, proposed by Zadeh an extension of fuzzy set theory and the other, taken by *Klir and Folger* as an extension dempster-shafer's theory of evidence.

2.8.1 Intuitive approach to Possibility Theory

Let S be a fuzzy set in a universe of discourse U which is characterized by its membership function $\mu_S(U)$ which is interpreted as the compatibility of $u \in U$ with the concept labeled S. Let X is a variable taking value in U and S act as a fuzzy restriction, R(X), associated with X. Then the proposition "X is S", which translates into R(X) = F associates a possibility distribution $\pi_X(x)$ with X which is postulated to be equal to R(X). The possibility distribution function $\pi_X(u)$ characterizing the possibility π_X is defined to be numerically equal to the membership function $\pi_S(u)$ of S that is $\pi_{(s)} = 0$ means that state *s* is rejected as impossible; $\pi(s) = 1$ means that state *s* is totally possible (= plausible).

2.8.2 The Axiomatic Approach to Possibility

For consonant body of evidence, the belief measure becomes necessity measure (η) and plausibility measure becomes possibility measure (π)

$$\text{Bel}(A \cap B) = \min[\text{Bel}(A), \text{Bel}(B)] \text{ becomes } = \eta(A \cap B) = \min[\eta(A), \eta(B)]$$

$$\text{Pl}(A \cup B) = \max[(\text{Pl}(A), \text{Pl}(B)] \text{ becomes } \pi(A \cup B) = \max[(\pi(A), \pi(B)]$$

$$\pi(A) = 1 - \pi(\overline{A}); \eta(A) = 1 - \eta(\overline{A})$$

Now if possibility distribution function is defined as:

$$r := U \to [0,1]$$

This mapping is related to possibility measure as: $\pi(A) = \max_{X \in A}(r(X))$ .Now the basic probability assignment is defined for consonant body of evidences as: $m = (\mu_1, \mu_2, \mu_3, \ldots\ldots\ldots, \mu_n)$ where $\sum_{i=1}^{n} \mu_i = 1$ and if possibility distribution function is defined as:

$$r = (\rho_1, \rho_2, \rho_3, \ldots\ldots\ldots, \rho_n) \text{ Where } \rho_i = \sum_{k=i}^{n} \mu_k = \sum_{k=i}^{n} m(A_k)$$

Possibility theory is a simple and versatile tool for modeling uncertainty. It is a unifying framework for modeling and merging linguistic knowledge and statistical data. Useful to account for missing information in reasoning tasks and risk analysis

3. Related Work

3.1 Rough set Theory

Rough set theory proposed by Pawlak [26][27], has become a well-established theory to resolve problems related to vagueness, uncertainty and incomplete information in variety of applications related to pattern recognition and machine learning. The main advantage of rough set theory in data analysis is that it does not need any preliminary or additional information about data like probability in statistics, or basic probability assignment in Dempster-Shafer theory, grade of membership or the value of possibility in fuzzy set theory.

3.2 Information and Decision Tables

Data are presented as a table, column of which are labelled by attributes, rows by objects of interest and entries of the table are attribute values [28][29]. Such tables are known as information systems, attribute-value tables, data tables or information tables. It can be represented as a pair $IS = (U, A)$ where $U = \{x_1, x_2, \ldots\ldots, x_n\}$ $U = \{x_1, x_2, \ldots\ldots, x_n\}$ is a non empty finite set of objects called the universe and $A = \{a_1, a_2, \ldots\ldots, a_n\}$ is a non- empty finite set of attributes such as a $U \rightarrow V_a$. $a \in A$ where V_a is value set of a. If V_a contains missing values for at least one attribute, then S is called an incomplete information system, otherwise it is complete. If the set of attributes A is divided into the set of condition attributes C and the set of decision attributes D such $A = C \bigcup D$ and $C \bigcap D$ is empty, the information system is called a decision table.

Attributes			
Id	**Headache**	**Nausea**	**Temperature**
1	no	yes	high
2	yes	no	very high
3	yes	yes	high
4	no	yes	high
5	yes	no	Normal
6	no	yes	Normal

Table 1. Example of Information table

Attributes				Decision
Id	**Headache**	**Nausea**	**Temperature**	**Flu?**
p1	no	yes	high	yes
p2	yes	no	high	yes
p3	yes	yes	very high	yes
p4	no	yes	normal	no
p5	yes	no	high	no
p6	no	yes	very high	yes

Table 2. Example of decision table with uncertainty

Rough set concept can be defined quite generally by means of topological operations, interior and closure, called approximations. Suppose we are given a set of objects U called the universe and an indiscernibility relation $R \subseteq U \times U$, representing our lack of knowledge about elements of U. For the sake of simplicity we assume that R is an equivalence relation [30]. Let X is a subset of U. We want to characterize the set X with respect to R. Basic concepts of rough set theory given below.

- The lower approximation of a set X with respect to R is the set of all objects, which can be for certain classified as X with respect to R (are certainly X with respect to R).

- The upper approximation of a set X with respect to R is the set of all objects which can be possibly classified as X with respect to R (are possibly X in view of R).
- The boundary region of a set X with respect to R is the set of all objects, which can be classified neither as X nor as not-X with respect to R.
- Set X is crisp (exact with respect to R), if the boundary region of X is empty
- Set X is rough (inexact with respect to R), if the boundary region of X is nonempty.

Let R is an indiscernibility relation or equivalence relation among indiscernible values.$[X]_R$ denotes the equivalence class of R containing x [31]. The elementary sets are the equivalence classes of R. Lower and upper approximations of a set X over the elements of the universe U are respectively

R-lower approximation of X

$$\underline{RX} = \{x \in \bigcup \mid [x]_R \subseteq X\}$$

R-upper approximation of X

$$\overline{RX} = \{x \in \bigcup \mid [x]_R \cap X \neq \varnothing\}$$

R-boundary region of X

$$R_{NR}(X) = \underline{RX} - \overline{RX}$$

Accuracy of Approximations given by

$$\alpha_R(X) = \frac{|\overline{R}(X)|}{|\underline{R}(X)|}[0 \leq \alpha_R \leq 1]$$

Where |X| denotes the cardinality of X. If $\alpha_R(X) = 1$, X is crisp with respect to R (X is precise with respect to R), and otherwise, if $\alpha_R(X) < 1$, X is rough with respect to R (X is vague} with respect to R).

3.3 Dependency attributes

Let C and D be subsets of A. We say that D depends on C in a degree $k(0 \leq k \leq 1)$ denoted by $C \rightarrow_k D$ such that

$$k = \gamma(C, D) = \frac{|POS_c(D)|}{|U|}$$

Where $POS_c(D)$ is called C-positive region of D. The coefficient k is called the degree of the dependency; if k = 1 we say that D depends totally on C, and if k < 1 we say that D depends partially (in a degree k) on C. In other words, the coefficient k expresses the ratio of all elements of the universe, which can be properly classified to blocks of the partition U/D, employing attributes C [33].

3.4. Example

Consider the table 1 example the lower approximation of the set of patients having flu is the set {p1, p3, p6} and the upper approximation of this set is the set {p1, p2, p3, p5, p6}, whereas the boundary-line cases are patients p2 and p5. Similarly p4 does not have flu and p2, p5 cannot be excluded as having flu, thus the lower approximation of this concept is the set {p4} whereas - the upper approximation – is the set {p2, p4, p5} and the boundary region of the concept "not flu" is the set {p2, p5}, the same as in the previous case. Accuracy of approximation can be calculated $\alpha_R(\text{flu}) = 3/5$. It means that the concept "flu" can be characterized partially employing symptoms Headache, Muscle-pain and temperature. Dependency of attributes for {Headache, Nausea, Temperature}⇒{Flu}.weget k = 4/6 = 2/3, because four out of six patients can be uniquely classified as having flu or not, employing attributes Headache, Nausea and Temperature.

3.5 Reduction of attributes (Reduct and Core attributes)

The main idea of reduction of attributes or reduct in rough sets is to find a minimal subset of relevant attributes that have all the essential information of the data set, thus the minimal subset of the attributes can be used instead of the entire attributes set for rule discovery and to make correct decision or classification. The concept of a reduct can be defined as minimal subset R of the initial attribute set C such that for such that for a given set of attributes D, $\gamma_R(D) = \gamma_C(D)$. R is a minimal subset if $\gamma_{R-\{a\}}(D) \neq \gamma_R(D), \forall a \in R$. This means that any attribute removed from the subset will affect the dependency degree [34]. Hence a minimal subset by this definition may not be the global minimum (a reduct of smallest cardinality). A given dataset may have many reduct sets, and the collection of all reducts is denoted by [35][37]

$$R_{all} = \{X \mid X \subseteq C, \gamma_X(D) = \gamma_C(D); \gamma_{X-\{a\}}(D) \neq \gamma_X(D), \forall a \in D\}$$

The intersection of all the sets in R_{all} is called the core, denoted by $CORE(C) = \bigcap RED(C)$,where RED(C) is the set of all reducts of C. For example there are two relative reducts with respect to Flu, {Headache, Temperature} and {Nausea, Temperature} of the set of condition attributes {Headache, Muscle-pain, Temperature}. That means that either the attribute Headache or Nausea can be eliminated from the table or consequently instead of table 3 we can use either

Patient	Headache	Temperature	Flu
p1	no	high	yes
p2	yes	high	yes
p3	yes	very high	yes
p4	no	normal	no
p5	yes	high	no
p6	no	very high	yes

Table 3

Patient	Nausea	Temperature	Flu
p1	yes	high	yes
p2	no	high	yes
p3	yes	very high	yes
p4	yes	normal	no
p5	no	high	no
p6	yes	very high	yes

Table 4

For table 4 the relative core of with respect to the set {Headache, Nausea, and Temperature} is the Temperature. This confirms our previous considerations showing that Temperature is the only symptom that enables, at least, partial diagnosis of patients. Using the concept of a value reduct, table 3 and table 4 can be simplified as follows

Patient	Headache	Temperature	Flu
p1	no	high	yes
p2	yes	high	yes
p3	-	very high	yes
p4	-	normal	no
p5	yes	high	no
p6	-	very high	yes

Table 5

Patient	Nausea	Temperature	Flu
p1	yes	high	yes
p2	no	high	yes
p3	-	very high	yes
p4	-	normal	no
p5	no	high	no
p6	-	very high	yes

Table 6

3.6 Rule Generation

The following rules can be generated from the table 4 we get the following rules
if (Headache, no) and (Temperature, high) then (Flu, yes),
if (Headache, yes) and (Temperature, high) then (Flu, yes),
if (Temperature, very high) then (Flu, yes),
if (Temperature, normal) then (Flu, no),
if (Headache, yes) and (Temperature, high) then (Flu, no),
if (Temperature, very high) then (Flu, yes).
From table 5 we get the following rules

4.7 Problems with Rough Set Theory

- All tuples are treated with equal importance
- All training examples must be crisply classified
- The lower and upper approximations of a concept are defined based on the strict set inclusion operation

In order to improve the accuracy within the rules the following parameters have to be tested is tested i.e. A decision rule f_x is included in U is consistent or deterministic if for every y included in U, $y \neq x$ (f_x/Cond=f_y /Cond).(f_x/Dec=f_y /Dec); otherwise the decision rule f_x is nondeterministic or inconsistent. In a similar manner, a decision table S is deterministic if all of its decision rules are deterministic; other-wise the decision table S is nondeterministic [37][38]

Attributes				Decision
Id	Temperature	Nausea	Headache	Flu?
p1	very high	yes	no	yes
p2	high	yes	no	no
p3	very high	yes	no	yes
p4	normal	no	no	no
p5	high	no	yes	yes
p6	very high	no	yes	yes

Let be R1 and R2 an example of decision rules from the decision table 7:

- R1: IF (Temperature = very high AND Nausea = yes AND Headache = no) then (Flu =Yes)
- R2: If (Temperature =high) then (Flu =Yes) OR (Flu =No)

A decision rule may be characterized by the most specific definitions from the above rules

- **Rule length:** no of conditional attributes in the given rule. As an example, the length of R1=3.
- **Rule strength:** is count of objects in the data set having the property described by the rule conditions and decisions. As an example, the rule strength of R1=4.
- **Exact rule:** the outcome of an exact rule corresponds to one or more different conditions. Exact rules are generated from the set of objects in the lower approximation. As an example, R1 is an exact Rule.
- **Approximate rule:** The same condition of an approximate rule corresponds to more than one outcome. Approximate rules are generated for the boundary. R2 is an example of approximate rule.
- **Rule coverage:** is the proportion of objects contained in the training set, identified by this rule. As an example, the rule coverage of R1=3/4=0.75
- **Rule acceptance:** the rule acceptance measure may be expressed as the count of condition terms of a rule. It is a subjective measure that reflects the confidence of the user in the extracted rules. It is a generalization of the rule support and rule coverage.
- **Discrimination level (DL):** DL measures the level of precision of a rule that represents the corresponding objects.
- **Decision support measure (DSM):** is the total number of rules that support a decision. A DSM may be expressed by the number of objects from the training set supporting the decision.
- **Decision redundancy factor (DRF):** is the count of mutually exclusive feature sets related to the same decision.

For more accuracy in the decision rules we have to calculate the following

Let us assume that

- $R = (R, A \cup \{d\})$ is a universal decision table,
- $A = (U, A \cup \{d\}$ is a given decision table $(U \subseteq R)$,
- $\rho_t \in R$ is tested object,
- $Rul(X_j)$ is a set of all calculated basic decisions for A ,classifying objects of the decision class X_j,
- $MRul(X_j, \rho_t) \subseteq Rul(X_j)$ is a set of all decisions rules for $Rul(X_j)$ matching tested objects ρ_t .

Several rules can be defined for the rule set $MRul(X_j, \rho_t)$ depending on the number of rules from this set matching tested object, the number of objects supporting decision rules from this set and the stability coefficient of rules.

The simple strength of a decision rule set is defined by

$$SimpleStregth(X_j, \rho_t) = \frac{card(MRul((X_j, \rho_t))}{card(MRul((X_j))}$$

The maximal strength of a decision rules set is defined by

$$MaximalStrength(X_j, \rho_t) = \max_{r \in MRul(X_j, \rho_t)} \left\{ \frac{Supp_A(r)}{card(|d = v_d^j|_A)} \right\}$$

The basic strength of a decision rule set is defined by

$$BasicStrength(X_j, \rho_t) = \frac{\sum_{r \in MRul(X_j, \rho_t)} Supp_A(r)}{\sum_{r \in Rul(X_j)} Supp_A(r)}$$

The global strength of a decision rule set is defined by

$$GlobalStrength(X_j, \rho_t) = \frac{card\left(\bigcup_{r \in MRul(X_j, \rho_t)} |pred(r)|_A \cap |d = v_d^j|_A \right)}{card(|d = v_d^j|_A)}$$

The Stability Strength of a decision rule set is defined by

$$StabilityStrength(X_j, \rho_t) = \max_{r \in MRul(X_j, \rho_t)} \left\{ SC_A^{P(A)}(r) \right\}$$

The Maximal Stability Strength of a decision rule set is defined by

$$MaximalStabilityStrength(X_j, \rho_t) = \max_{r \in MRul(X_j, \rho_t)} \left\{ \frac{Supp_A(r)}{card(|d = v_d^j|_A)} \right\} \left\{ SC_A^{P(A)}(r) \right\}$$

The Basic Stability Strength of a decision rule set is defined by

$$BasicStabilityStrength(X_j, \rho_t) = \max_{r \in MRul(X_j, \rho_t)} \frac{\sum_{r \in MRul(X_j, \rho_t)} Supp_A(r).\, SC_A^{P(A)}(r)}{\sum_{r \in Rul(X_j)} Supp_A(r).\, SC_A^{P(A)}(r)}$$

These measures can be applied in constructing classification algorithms and to develop more efficient rule based to eliminate uncertainties in the given rules. The accuracy of the classification can be measured based on percentage of successful classification (PSC) on the dataset

$$PSC = \frac{\text{Number of Correctly Classified Instances}}{\text{Total No of Instances in Data Set}} \times 100$$

4. Proposed Method Classification Algorithm

Step1: Conditional attribute and decision attribute value provide positive integer value
Step2: Convert data positive integer form.
Step3: Construction of elementary sets using Rough set theory
Step4: Use discretization methods if the variables are continuous
Step5: Find core and dynamic reduct for attribute.
Step6: Find core and dynamic reduct for attribute values.
Step7: Create reduction table of subspace reduct.
Step8: Generate classification model based decision rules.
Step9: Selection of rules matching new case
Step10: Calculation of strength of the selected rule sets for any decision class (from above section)
Step11: Selection of decision class with maximal strength of the selected rule set
Step12: Predicted decision value for new case
The planned method was tested on several datasets from UCI Machine Learning Repository dataset [39]

Data Set	Data Types	Instance Sizes	Attribute Sizes	Training set Sizes	Test set Sizes	Rules Generated Sizes	Accuracy (PSC)
Breast Cancer	Categorical	286	10	500	149	17	97.32%
Mushroom	Categorical	8124	23	450	150	18	92.67%
Soyabean-Small	Categorical	307	35	625	625	80	91.7%
Glass	Numeric	214	10	144	70	13	85.71%
Iris	Numeric	150	4	150	150	18	93.5%
New-Thyroid	Numeric	215	6	215	215	49	86.5%

5. CONCLUSION

The need for managing uncertainty increases as uncertainty leads to incomplete information and unpredictability. There are many techniques developed to analyze the uncertainty. All the approaches have their weakness and strength. No one method or approach is ideal for any expert systems. It all depends upon the problem to be solved. Rough set theory mainly deals with methods to classify imprecise, uncertain, and incomplete information or knowledge expressed in terms of data acquired from experience. By converting decision table the rules into based on testing data set, attribute reduction can be used to select the most important rules. By calculating different rule characteristics we can find more accurate rules for accurate classification

REFERNCES

[1] Klir, G. and M. Wierman. Uncertainty-Based Information: Elements of Generalized Information Theory, 2nd edition. Physica-Verlag. 168pp, 1999
[2] M. A. Lee. and M. H. Smith. (1995), "Handling Uncertainty in Finance Applications Using Soft Computing," Proc. 3 rd Int'l Symp Uncertainty Modelling and Analysis.
[3] Anuradha, Kumar Sachin, Kumar Anubhav "Analysis of Information Uncertainty Management in Expert Systems" International Journal ofAdvanced Research in Computer Science and Software Engineering 3(9), September - 2013, pp. 899-903
[4] U. Fayyad, R Uthurusamy. "Data Mining and Knowledge Discovery in Databases", Communications of the ACM.Vol.39, No11, November 1996.
[5] C. Aggarwal and P. Yu, "A Survey of Uncertain Data Algorithms and Applications," IEEE Trans. Knowledge Data Eng., vol. 21, no. 5,pp. 609-623, May 2009.
[6] M. Chau, R. Cheng, B. Kao, and J. Ng, "Uncertain data mining: An example in clustering location data,"in PAKDD, ser. Lecture Notes in Computer Science, vol. 3918. Singapore: Springer, 9–12 Apr. 2006, pp. 199–204
[7] C.C. Aggarwal, "On Density Based Transforms for Uncertain Data Mining," Proc. Int'l Conf. Data Eng. (ICDE), pp. 866-875, Apr. 2007.
[8] Managing and Mining Uncertain Data, C. Aggarwal, ed. Springer,2009
[9] Aggarwal, C.C., Yu, P.S: A survey of Uncertain Data Algorithms and Applications. IEEETransactions on Knowledge and Data Engineering 21(5), 609–623 (2009)
[10] York, J., Madigan, D., Heuch, I. and Lie, R. T. (1995) Birth defects registered by double sampling: a Bayesian approach incorporating covariates and model uncertainty. Appl. Statist., 44, 227-242.
[11] A. Kolmogorov, Foundations of the Theory of Probability.Chelsea Pub., 1950.
[12] Martin. Basil., Christos. Papadopoulos., Donald. Sutherland. and Hoi. Yeung. (2001) "Application of Probabilistic Uncertainty Methods (Monte Carlo Simulation) In Flow Measurement Uncertainty Estimation," Int'l Conf.Flow Measurement.
[13] Dwlvedi Ashutosh, Mishra Deepak & Kalra Prem Kumar,(2006),"Handling Uncertainties-using Probability Theory to PossibilityTheory", IIT Kanpur, pp.1-12.
[14] Dubois Didier,(2006) "Possibility Theory and Statistical Reasoning"Computational Statistics & Data Analysis, 51(1):pp.47-69
[15] Shafer Glenn,(1976),"Mathematical Theory of Evidence", Princeton, N.J.: Princeton University Press, Business & economics,V42
[16] D. Dubois and H. Prade,(1992)"when upper probabilities are possibility measures",Fuzzy Sets and Systems Journal,49, 65-74
[17] R. Cheng, D. Kalashnikov, and S. Prabhakar, "Evaluating probabilistic queries over imprecise data," in ACM SIGMOD International Conference on Management of Data , 2003, pp. 551–562
[18] Denœux Thierry,(2012)"Belief functions:theory and applications", in 2nd International Conference on Belief functions, springer, France9
[19] Bernecker, T., Kriegel, H.P., Renz, M., Verhein, F., Zuefle, A.: Probabilistic frequent itemset mining in uncertain databases. In: Proc. of SIGKDD 2009, pp.119–128 (2009)
[20] L.A. Zadeh, "Fuzzy Sets as a Basis for a Theory of Possibility,"J. Fuzzy Sets and Systems,vol. 1, pp. 3-28, 1978
[21] M. Vazirgiannis, M. Halkidi. "Uncertainty handling in the datamining process with fuzzy logic", in the proceedings of the IEEE-FUZZY conference, San Antonio, Texas, May, 2000
[22] L.A. Zadeh, "Fuzzy Sets,"J. Information and Control,vol. 8, pp. 338-353, 1965
[23] S. Chiu. "Extracting Fuzzy Rules from Data for Function Approximation and Pattern Classification".Fuzzy Information Engineering- A Guided Tour of Applications. (Eds.: D. Dubois, H. Prade, R.Yager), 1997
[24] L.A. Zadeh, "Fuzzy Logic,"IEEE Computer Magazine,vol. 21, no. 4,pp. 83-93, Apr. 1988.
[25] F. Dweiri and M. Kablan, "Using Fuzzy Decision Making for the Evaluation of the Project Management Internal Efficiency,"Decision Support Systems, vol. 42, pp. 712-726, 2006.

[26] Z. Pawlak, "Rough Sets," *International Journal of Computer and Information Sciences*, vol. 11, pp. 341-355, 1982.

[27] Grzymala-Busse, J.W. (1988) 'Knowledge acquisition under uncertainty a rough set approach', *Journal of Intelligent & Robotic Systems*, Vol. 1, No. 1, pp.3–16.

[28] Pawlak, Z., Grazymala-Busse, J.W., Slowinski, R. and Ziarko, W. (1995) 'Rough sets', Communications of the ACM, Vol. 38, pp.88–95.

[29] J. Li, "Rough set based rule evaluations and their applications." Waterloo, Ontario: University of Waterloo, 2007, pp. 191.

[30] Gordon, A. D. (1981) Classification: Methodsfor the Exploratory Analysis of Multivariate Data. London: Chapman and Hall.

[31] Taylor, B. N. and C. E. Kuyatt. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. Technical report, National Institute of Standards and Technology Technical Note 1297, Gaithersburg, MD, 1993

[32] S. Wasserkrug., A. Gal. and O. Etzion. (2005) "A Model for Reasoning with Uncertain Rules in Event Composition Systems," Proc. 21st Conf. Uncertainty in Artificial Intelligence (UAI '05), pp. 599-606

[33] S. Tsumoto, "Accuracy and coverage in rough set rule induction," in *Rough Sets and Current Trends in Computing*, J. J. Alpigini, et. al., Ed. Berlin, Heidelberg: Springer-Verlag, 2002, pp. 373-380

[34] J. Li and N. Cercone, "Introducing a rule importancemeasure," in *Lecture Notes in Computer Science*: Springer, 2006, pp. 167.

[35] Grzymala-Busse, J.W. (1997) 'A new version of the rule induction system LERS', *Fundamenta Informaticae*, Vol. 31, No. 1,pp.27–39

[36] Sai, Y., Nie, P., Xu, R. and Huang, J. (2006) 'A rough set approach to mining concise rules from inconsistent data', Proceedings of IEEE GRC 2006, IEEE, Atlanta USA, pp.333–336.

[37] Slowinski, R. (1992) In Intelligent Decision Support: Handbook of Applications and Advances of the Rough Set Theory, Kluwer Academic Publishers, Norwell, MA, USA.

[38] Li, Jiye. & Cercone, Nick., "Assigning Missing Attribute Values Based on Rough Sets Theory", IEEE International Conference on Granular Computing, Pp. 607 – 610, May, 2006

[39] A. Asuncion and D. Newman, UCI Machine Learning Repository, http://www.ics.uci.edu /mlearn/ MLRepository .html, 2007