International Journal of Soft Computing 13 (6): 149-157, 2018
ISSN: 1816-9503

Extracting Knowledge from Incomplete Dataset by Tolerance Rough Sets with Consistency Measure

[1]Elsayed Radwan and [2]Y. Alkazemi Basem
[1]College of Computer and Information System, Umm Al-Qura University, Makkah, Kingdom of Saudi Arabia (KSA)
[2]Faculty of Computer and Information Sciences, Mansoura University, Mansoura, Egypt

Abstract: Knowledge extraction from various datasets is suffering from missing attribute values, representing the non-categorical ones, besides the inconsistency among the local rules. These shorts reflect two main problems, information granules representation and superfluous feature existence. This study treats these problems based on tolerance rough sets. A similarity relation is considered based on the matching percentage among objects in order to decrease the knowledge granularity and hence, increase the information capacity. In order to measure the dependencies among features, a tolerance measure based on step-deck coordination function is defined as vague inclusion function. For feature extraction process, a consistency function is defined. It is proved that the consistency measure promise an increase in classification accuracy. More specifically, the proposed model receives a base method of tolerance rough set to generate local rules that describe the application algorithm. The details and limitations of the proposed model are discussed where a comparative study is appeared.

Key words: Knowledge extraction, similarity relation, tolerance rough set, step-deck coordination function, consistency measure, classification accuracy

INTRODUCTION

Data mining becomes a vital field of science now a days as it can help inferring new interesting facts from data at hand and classify features accordingly. However with the increasing amount and types of data received from IoT devices or WSN nodes for instance, it might be hard to identify the appropriate classes for a set of extracted features (knowledge) that they represent as the level of uncertainty for deciding the appropriate class for a given data set can be high (Wang *et al*., 2013; Sun *et al*., 2014; Wu *et al*., 2009, 2017)

The knowledge extraction problem is fundamentally enlarged because human knowledge is mainly incomplete, inexact and approximate. For an accurate decision output, the decision rules should be robust for data variation. Since, robust output should prevent the variation in the dataset with a predefined threshold, it is important to automate the knowledge extraction process based on permission factor (tolerance). Based on the suitable feature selection mechanism that interact with incomplete decision system and rule evolving algorithm, accurate and robust outputs can be achieved. Knowledge extraction from incomplete dataset can be more challenging as it requires deciding upon the appropriate features to consider in order to identify their class of knowledge, hence, needs a considerable human effort to handle and manage that. Thus, improving the rule induction process is an emergent task and poses a major research challenge. Unfortunately, most datasets gathered from different sources appears to be incomplete, irrelevant, redundant, imprecise and high features dimensionality. Consequently, the knowledge extracted might be unclassifiable at all or maybe classified incorrectly under different irrelevant type of knowledge. Therefore, modeling granularity and feature reduction of Incomplete Information System (IIS), uncertain knowledge base, provide a fruitful area for researchers about knowledge representation (Wu and Wang, 2008; Ali *et al*., 2015; Jusoh *et al*., 2007).

According to the previous research, many researchers focused on developing systems that handle the dataset problems individually, domain specific problem. A generalized case depends on soft computing techniques. By Rivero *et al*. (2005), a mathematical tool, find a subset of features that predicate the decision concept as the same original features but not the optimal one that struggle disturbance and noise data (Ali *et al*.,

Corresponding Author: Elsayed Radwan, Department of Scientific Research, Umm AL-Quran University, Makkah, Kingdom of Saudi Arabia (KSA)

2015). However, Pawlak rough sets model is restricted with all dataset that should be discrete and static. Some extensions of rough sets are presented like covering rough sets model, dominance-based rough sets, variable precision rough sets and tolerance rough sets (Huang *et al*., 2017; Ramanna, 2011; Dai *et al*., 2014; Azar *et al*., 2017). These extensions need a prior knowledge about the data sets except the regular rough sets model. Also, they depends on defining tolerance relation instead of equivalence relation which result significant slowness in feature extraction. Moreover, the machine learning algorithms like C4.5 are used in acquiring knowledge in incomplete decision system (Ibrahim *et al*., 2011) wherever it still suffer from the irrelevant feature problem. Also, in incomplete decision systems, the machine learning algorithms are still restricted with non-incremental uncertainty measure problem (Sun *et al*., 2012). In addition, complicated mathematical function for measuring uncertainty should be defined. By this study, tolerance rough set model with a partial similarity relation and step-deck coordination function (vague inclusion function) is introduced to handle these problems.

As a result, extracting and reducing knowledge should have innovative heuristics technique to reduce the time complexity and achieve robust classification with minor disturbance. The main purpose of this study is to reduce the number of granules by means of partial similarity relation that depends on the matching percentage among objects. Also, increase the information capacity based on step-deck coordination function which degrade the uncertain knowledge since small variation of values do not result a large changes in vague inclusion function, feature dependency function. This study presents an innovative method based on the tolerance rough set (Mac Parthalain and Shen, 2009; Hu, 2015) and consistency measure with step-deck coordination function to extract the core feature and hence, discover the admissible rules . Tolerance rough set model employs a consistency degree to measure the certainty information. The information gained by the consistency measure guide the algorithm of feature reduction process. Now regarding to tolerance rough set, it is right that the learned rules are quite precise. The evolved rules by TRSM+Consistency measure are enhanced and the classification accuracy is increased. In contrast with other previous of machine learning algorithms, the experiment and the result are declared in the UCI repository where a comparative study is done.

Tolerance rough sets: Ali *et al*. (2015) and Rivero *et al*. (2005), introduced by Pawlak is a new mathematical theory to imprecision, vagueness and uncertainty. In rough set approach, an approximation space is defined for any arbitrary subset of the universe of discourse, $X \subseteq U$. Rough Set Model (RSM) generates two subsets by means of an equivalence relation E; Namely, lower and upper approximations:

$$L\big(E(X)\big) = \{x \in U : [X]_E \subseteq X\} \quad A\big(E(X)\big) = \{x \in U : [X]_E \cap X \neq \varnothing\}$$

Through the lower and upper approximations, one can not only extract the decision rules that are hidden in the database but also select the minimal subset of data that predicate the same meaning as the whole set. The Pawlak approximation space (U, E) measures the approximation accuracy by:

$$\alpha_E(XX) = |L\big(E(X)\big)| / |A\big(E(X)\big)|$$

where |.| defined the cardinality measure. RSM reduces the data efficiently if a 1-1 correspondence indiscernibility relation is defined which is frequently happen.

A tuple DS = $(U, C \cup D, V, \rho)$ is called a decision system where U is a non-empty set called the universe of discourse, C and D, $C \cap D = \phi$ are the condition and decision attributes respectively. For each attribute $a \in C \cup D$ there is a function $\rho: U \times a \rightarrow V_a$ where V_a is called the value domain of attribute, a and $V = U_{a\in C\cup D} V_a$. The main idea of any DS is to extract the knowledge, like if-then rules by reducing the condition set of attributes and remove the superfluous attribute values which cause a missing value without effecting the semantics of DS.

If for any $a \in C$ there is a missing attribute value, i.e., $* \in V_a$ or a nominal value, the DS is called Incomplete Decision System, IDS. In the recent application the missing attribute value is considered as any possible feature value. Thus, the equivalence relation can't construct a partition of the universe U. By defending a similarity relation or a tolerance relation instead of equivalence relation to solve the problem of missing attribute values, overlapping granules (classes) is generated:

$$TR(B) = \{(x,y) \mid \rho(x,a) = \rho(y,a) \text{ or } (x,a) = * \text{ or } \rho(y,a) = *,\ a \in B \text{ and } x,y \in U\}$$

The approximation space that is result by tolerance relation is called tolerance space (Ho and Nguyen, 2002) $R = (U, I, \upsilon)$ where $I: U \rightarrow P(U)$, P(U) is the power set of U is called the uncertain function. The uncertain function maps every element $u_i \in U$ to its corresponding class generated by the tolerance (or similarity) relation. This function should be reflexive and symmetric.υ: $P(U)\times P(U)\rightarrow[0, 1]$ is the vague inclusion function that measures the degree of inclusion of such subset. This function should satisfy the monotonicity property with respect to the second argument. By tolerance rough set model the uncertainty should be reduced and the local rules are discovered.

MATERIALS AND METHODS

Feature reduction based on consistency measure with tolerance: To improve the prediction accuracy in incomplete decision system irrespective to domain size dimensionality and information granules distribution, a feature subset reduction and tolerance permission altogether is needed to overcome the existing issues of classification systems. Structurally, this model involves two phases. First, the optimal number of information granules that construct the base IF-Then rules. The induction algorithm should be able to handle the missing attribute value in the decision system in addition to the nominal ones. Second, the induction algorithm should select the minimal set of significant features that build the whole features semantics. This model tackle the interpretability of the rule algorithm that receive the local rules from different experts and evolve them altogether to filter out informative rules.

Reducing granulates by partial similarity relation: The Pawlak approximation space (Rivero *et al.*, 2005) depends totally on defining a one-one correspondence relation on the universe of discourse to exploit the reduced set of features. In feature reduction there is no way to automatically compute the representative classes (information granules) by means of rough sets. Since, rough sets model requires a transitive property to handle disturbance and noise data, missing and continuous attribute values, a modified rough sets model for knowledge retrieval is defined to treat tolerance classes instead of equivalence classes. Tolerance Rough Set Model (TRSM) based on non-hierarchical cluster algorithm can be applied in large dataset with missing attribute values (Ho and Nguyen, 2002). In purpose of constructing tolerance classes a similarity relation is needed. Similarity relation denotes ordered pairs among objects that are possibly indiscernible in terms of attribute values taking into account the case of missing values. In this study, a similarity relation is defined based on matching percentage between two attribute values to be equal or, in case of continuous values, their intersection is hold by at least cut as depicted in Eq. 1:

$$\text{Match}_{\varsigma}(x, y, a) = \begin{cases} 1, \rho(x, a) = \rho(y, a) \\ ,\text{or} \dfrac{|\rho(y, a) \cap \rho(x, a)|}{\max|\rho(x, a)|, |\rho(y, a)|} \geq \end{cases}$$
$$\varsigma, \text{or}\, \rho(y, a) = *, \text{or}\, \rho(x, a) = 0 \quad (1)$$

where x, y$\in$U, $\varsigma\in(0.5, 1]$ if $\varsigma = 1$ then both values should be equal otherwise the two set values should intersect by minimum interval length by at least ς cut value. The match relation is a tolerance relation, reflexive and symmetric relation (Dai and Tian, 2013). The match relation on the set $X\subseteq U$ is illustrated in. Algorithm 1 the matching relation on a set X and an attribute a.

Algorithm 1; Match (X, a):

```
Input: X⊆U X⊆U, a∈A
Output:{X1, X2, ... , Xm} ⊂ P(X)
1. Let Z = {z|z = ρ(X, a)}; S1 {X∈U = ρ(x, a) = *}, S2 = XX-S1
2. Z1 = Z. Distinct-{*}, i =1
3. Foreach (z in Z1)
      a- Xi = φ
      b- fpreach (x in S2)
            If (mach(ρ(x, a), z)          Xi = XiUx
      c- Xi = XiUS1
      d-i++;
4. return {X1, X2, ... , Xm}, m = |Z1|
```

Consider IIS = (U, C) be an incomplete information system for any set of attributes $P\subseteq C$ the similarity relation of P on $U\times U$ is defined as follow:

$$\text{Sim}_{P,\alpha}(U) = \{(x, y) \in U\times U \frac{\sum_{i=1}^{\|p|} \text{Match}_{\varsigma}(x, y, a_i)}{|P|} \geq \quad (2)$$
$$\alpha, a_i \in P, i = \{1, \dots, \|\}$$

In Eq. 2 a partial similarity relation is defined which is reflex and symmetric. Thus, the tolerance classes corresponding to Eq. 2 the minimal description is given by:

$$S_{P,\alpha}(X) = \{y \in U : (x, y) \in \text{Sim}_{P,\alpha}(U)\} \quad (3)$$

The set $S_{P,\alpha}(x)$ (minimal description) should construct a cover set, i.e., $S_{P,\alpha}(x)\neq\varnothing$ and $\cup S_{P,\alpha}(x) = U$. Thus, the indiscernibility relation on the universe U is defined as:

$$U \,|\, \text{Sim}_{P,\alpha} = \{\text{Sim}_{P,\alpha}(x), \alpha \in (0.6, 1], x \in U\} \quad (4)$$

To be able to compute the $S_{P,\alpha}(x)$, illustrated as algorithm 2, the maximal description set should be first defined where:

$$\max_{desc}(x) = \{K \subset U \,|\, x \in K \wedge K \in \cap_{ai\in P} \text{Match}(U, a_i)\}$$

Algorithm 2; The algorithm for computing the minimal description set, maximal description set and superflous attributes:

Algorithm 2: $S_{P,\alpha}(x)$ and the superfluous set of attributr S:

Input $X\subseteq U$, $P\subseteq A$, $\alpha = 1$
Output: The minimal description $S_{P,\alpha}(x)$, $x\in X$ and Max-Desc = $\{Y_1, \ldots, Y_{|U|}\}$, S
Pre: Rearrange the set of attributes with minimum number of missing values P
1. Let Max-Desc = Match (X, a_0), $a_0\in$ P
2. Let T = $\{a_0\}$, S = ϕ
3. foreach ($a_i\in$ P-T)
   a- T = TU$\{a_i\}$
   b- TempCor = ϕ, bool = true
   c- foreach (Y in Max-Desc)
      i- Temp = Match (Y, a_i)
      ii- foreach (y_j in Temp)
         if ($y_j\notin$ TempCor)
            TempCor = TempCor$\cup\{y_j\}$
         bool = false
   d- Max-Desc = ($Max_{Desc}\cup$TempCor).distinct
   e- if (bool)
      S = S$\cup\{a_i\}$
4. Fpreach (x in X)
   a- $S_{P,\alpha}(x) = \varphi$
   b- foreach (Y in Max-DDesc)
      If (x in Y) $S_{P,\alpha}(x) = S_{P,\alpha}(x) \cup Y$
5. Return $S_{P,\alpha}(x)$, S

Hence, the approximation space based on the tolerance rough sets properties for any concept set Y, the lower and upper approximation is given by:

$$L(P_\alpha(Y)) = \cup\{S_{P,\alpha}(X): x\in U, \upsilon(S_{P,\alpha}(X), Y) > 1\text{-}\varepsilon\} \quad (5)$$

$$A(P_\alpha(Y)) = \cup\{S_{P,\alpha}(X): x\in U, \upsilon(S_{P,\alpha}(X), Y) > 1\text{-}\varepsilon\} \quad (6)$$

Where: $\nu\ 2^U\times 2^U\rightarrow[0, 1]$ is vague inclusion function:

$$\upsilon(X, Y) = \frac{|X\cap Y|}{|X|}$$

This function is clearly monotonous with respect to the second argument. The boundary region of Y based on the knowledge $P\subseteq C$ is given as:

$$\begin{aligned} BND(P_\alpha(Y)) &= A(P_\alpha(Y))\text{-}L(P_\alpha(Y)) \\ &= \cup\{S_{P,\alpha}(X): x\in U, \varepsilon < \upsilon(S_{P,\alpha}(X), Y) < 1\text{-}\varepsilon\} \end{aligned}$$

The negative region of Y based on the knowledge is given as:

$$NEG(P_\alpha(Y)) = \cup\{S_{P,\alpha}(X): x\in U, \upsilon(S_{P,\alpha}(X), \overline{Y}) > 1\text{-}\varepsilon\}$$

Then the positive region and the dependence measure are defined as follow:

$$POS_{P,\alpha}(D) = U_{Y\in U\|D} L(P_\alpha(Y)) \quad (7)$$

$$\gamma_{P,\alpha}(D) = \frac{|POS_{P,\alpha}(D)|}{|U|} \quad (8)$$

Feature selection based on consistency measure with step-deck relation: The main goal of this part is to utilize the information contained in the positive region, the region of certain information. This information serve in discovery of more compact feature subsets and increase the classification accuracy.

Lemma (1) (Herawan *et al.*, 2010): If $U|D = \{Y_1, Y_2, \ldots, Y_n\}$ is a equivalence relation on the universe of discourse U then $\sum_{i=1}^{n}\upsilon(X, Y_i) = 1$ for any $X\subseteq U$.

Proof: For any $X\subseteq U$ and $Y_j\in U|D$, it is known that $U_{j=1}^{n} Y_j = U$. Hence, $X\subset U_{j=1}^{n} Y_j = U$. There exist at least $Y_j\in U|D$ such that $\upsilon(X, Y_j)>0$, thus $U_{j=1}^{n}(Y_j\cap U) = X$., i.e., $\upsilon(X, U_{j=1}^{n} Y) = 1$. Since, $Y_j\in U|D$ for all j=$\{1, \ldots, n\}$ are equivalence classes $Y_j\cap Y_k = \varnothing$, $i\neq k$. Hence, $\upsilon(X, U_{j=1}^{n} Y_j) = \sum_{i=1}^{n}\upsilon(X, Y_i) = 1$.

Definition (1) (Sun *et al.*, 2012, 2014): Let (U, C, D) be incomplete decision system, $P\subseteq C$, $U = \{u_1, u_2, \ldots, u_{|U|}\}$, $U|D = \{Y_1, \ldots, Y_2\}$ the distribution of under the knowledge u_i P is:

$$\begin{aligned}\mu_P(u_i) &= (\upsilon(S_{P,\alpha}(u_i), Y_1), \upsilon(S_{P,\alpha}(u_i), Y_2), \ldots, \upsilon(S_{P,\alpha}(u_i), Y_n)) = \\ &(Y_1^{P,\alpha}(u_i), Y_2^{P,\alpha}(u_i), \ldots, Y_n^{P,\alpha}(u_i))\end{aligned}$$

Then, $P\subseteq C$ is a distribution reduct of the IDS if the following conditions are satisfied $\mu_P(u_i) = \mu_C$ for all $u_i\in U$. For any $Q\subset P$ there exist $u_j\in U$ such that $\mu_Q(u_j)\neq\mu_C(u_j)$. In IDS incomplete decision for each feature $a_k\in C$ and $U|D = \{Y_1, Y_2, \ldots, Y_n\}$ the vector $D_j^{a_k,\alpha} = (Y_j^{a_k,\alpha}(u_1), Y_j^{a_k,\alpha}(u_2), \ldots, Y_j^{a_k,\alpha}(u_{|U|}))$, where $Y_j^{a_k,\alpha}(u_1) = \upsilon(S_{a_k,I}(u_i), Y_j)$ represents the membership distribution of each object in U corresponding to the cluster Y_j. In other words, $D_j^{a_k,\alpha}$ represents the membership value of each object $u_i\in U$ that is included in the local rule corresponding to the decision class Y_j. This distribution can include the information in the coordination of the condition attribute set C to the clustering decision D. Moreover the decision accuracy can be reflected as a consistency measure. On behave of our approach, a coordination function is defined in purpose of using it to measure the consistent degree.

Definition 2 (Chen *et al.*, 2006): Let θ: $[0, 1]\rightarrow[0, 1]$ be a coordination function if it satisfies $-\theta(u) = 1$ represents an element belongs to the positive or negative regions $-\theta(u) = 1$, a symmetric function $-\theta(u)$ on [0, 0.5] monotonic decreasing function. This map indicate that if an object locate in the positive or negative region then it has value one but it takes the value 0 if it locates the object in the boundary region. Moreover, this map is symmetric ½ and 1/2. Finally, it specifies monotonic feature. In order to increase the information capacity and avoid the data disturbance, a step-deck coordination map $\Phi_\varepsilon(u) = 2\lfloor|u-0.5|+\varepsilon\rfloor$, as illustrated in Fig. 1 where $\varepsilon\in(0, 0.1)$ is the tolerance percentage is used.

The main objective of step-deck measure is to reduce the number granules according to the parameter ε, thus, the boundary region is reduced in accordance. Moreover, the positive and negative regions are monastically increased as result of is increased. By the step-deck coordination function computation time is also reduced.

Definition 3: consider IDS = (U, C, D) be incomplete decision system and $U|D = \{Y_1, Y_2, \dots, Y_n\}$ then the consistent degree of Y_j is:

$$CON_C(j) = \frac{1}{|C||U|}\sum_{k=1}^{|C|}\sum_{k=1}^{|U|}\Phi_\varepsilon\left(Y_j^{a_k,\alpha}(u_1)\right) \quad (9)$$

And the consistent degree of IDS as:

$$CON_C(D) = \frac{1}{n}\sum_{j=1}^{n} CON_C(j)$$

Proposition 1: consider IDS = (U, C, D) be incomplete decision system and P, $Q\subseteq C$. If $Sim_{P,\alpha}(U) = Sim_{Q,\alpha}(U)$ then $CON_P(D) = CON_Q(D)$.

Proof: Suppose P, $Q\subseteq C$ and $Sim_{P\alpha}(U) = Sim_{Q\alpha}(U)$ from the definition of tolerance relation, it follow that $S_{P\alpha}(u_i) = Sim_{Q\alpha}(u_i)$ for any $u\notin U$. Then we conclude that $\upsilon(S_{P,\alpha}(u_j)(Y_j)) = \upsilon(S_{Q,\alpha}(u_j)(Y_j))$, i.e., $Y_j^{P,\alpha}(u_j) = Y_j^{P,\alpha}(u_j)$ for all $Y_j\in U|D$. Since, the coordination map is monotonic $\Phi(Y_j^{P,\alpha}(u_i)) = \Phi(Y_j^{Q,\alpha}(u_i))$. Thus, $CON_P(j) = CON_Q(j)$. Hence, $CON_P(D) = CON_Q(D)$.

Proposition 2: Consider IDS = (U, C, D) be incomplete decision system and $P\subseteq Q\subseteq C$., if $CON_P(j) = CON_Q(j)$ for all $Y_j\in U|D$ then $Sim_{P,\alpha}(U) = Sim_{Q,\alpha}(U)$.

Proof: Since, $P\subseteq Q\subseteq C$ then we have P is coarser than Q. Thus, $S_{Q\alpha}(u_i) = Sim_{P\alpha}(u_i)$ and as a result of $CON_P(j) = CON_Q(j)$, thus, we have $\upsilon(S_{P,\alpha}(u_j)(Y_j)) = \upsilon(S_{Q,\alpha}(u_j)(Y_j))$ for any $u_I \in U$. Therefore, $S_{q\alpha}(u_i) = Sim_{P\alpha}(u_i)$. Hence, $Sim_{P\alpha}(U) = Sim_{Q\alpha}(U)$.

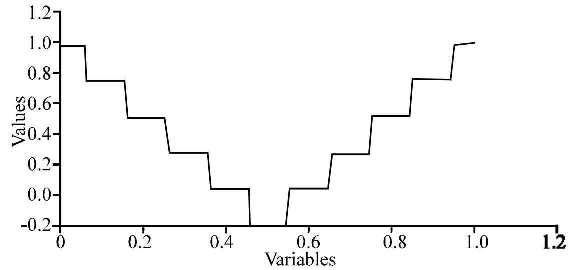

Fig. 1: The matching relation on a set X and an attribute a

Corollary (1): consider IDS = (U, C, D) be incomplete decision system and $P\subseteq C$., if $CON_P(j) = CON_Q(j)$ for all $Y_j\in U|D$ if and only $U|Sim_{P,\alpha} = U|Sim_{C,\alpha}$.

Proof: The proof comes directly from proposition (1) and proposition (2).

Proposition (3) (Monotonicity): Let IDS = (U, C, D) be incomplete decision system and P, $Q\subseteq C$. If P is finer than Q then $CON_P(j) = CON_Q(j)$ for all $Y_j\in U|D$.

Proof: Let $U|Sim_{P,\alpha} = \{S_{P,\alpha}(u_1), S_{P,\alpha}(u_2), \dots, S_{P,\alpha}(u_{|U|})\}$ and $U|Sim_{Q,\alpha} = \{S_{Q,\alpha}(u_1), S_{Q,\alpha}(u_2), \dots, S_{Q,\alpha}(u_{|U|})\}$. Since, P is finer than Q, then for all $u_i\in U$ such that $S_{P,\alpha}(u_i)\subseteq S_{Q,\alpha}(u_i)$ and there exist $u_i\in U$ satisfies $S_{P,\alpha}(u_i)\subseteq S_{Q,\alpha}(u_i)$. Therefore, $POS_{QP,\alpha}(Y_j)\subseteq POS_{P,\alpha}(u_j)$ for all $Y_j\in U|D$. Thus, $|POS_{QP,\alpha}(Y_j)|\subseteq |POS_{P,\alpha}(u_j)|$ for all $Y_j\in U|D$. From the definition of positive region $\sum_{i=}^{|U|}\upsilon(S_{Q,\alpha}(x),Y_j) < \sum_{i=}^{|U|}\upsilon(S_{P,\alpha}(x),Y_j)$. Since, the coordination map is monatomic, thus:

$$\sum_{i=1}^{|U|}\Phi\left(\upsilon\left(S_{Q,\alpha}(x),Y_j\right)\right) \leq \sum_{i=}^{|U|}\Phi\left(\upsilon\left(S_{P,\alpha}(x),Y_j\right)\right)$$

$$\sum_{i=1}^{|U|}\frac{1}{Q}\sum_{k=1}^{|Q|}\Phi\left(\upsilon\left(S_{a_k,\alpha}(x),Y_j\right)\right) \leq \sum_{i=1}^{|U|}\frac{1}{P}\sum_{k=1}^{|P|}\Phi\left(\upsilon\left(S_{a_k,\alpha}(x),Y_j\right)\right)$$

Thus, $CON_P(j)\geq CON_Q(j)$ for all $Y_j\in U|D$. According to, the monotonicity proposition, without considering the case $U|D = \{U\}$, the finer the knowledge is the lager the $CON_{P(D)}$ should be. Consistency measure can serve to choose the optimal set of features that preserving the semantics of the features.

Definition (IV): Let IDS = (U, C, D) be incomplete decision system, $P\subseteq C$ and $U = \{u_1, u_2, \dots, u_{|U|}\}$, $U = \{Y_1, \dots, Y_n\}$ the set P is said to reduct of the IDS if and only if $CON_P(j)\geq CON_C(j)$ for all $Y_j\notin U|D$ and for only any $Q\subset P$ there exists $Y_k\in U|D$ such that $CON_P(j)\geq CON_C(j)$.

Proposition 4: Let IDS = (U, C, D) be incomplete decision system, $P \subseteq C$ and $U|D = \{Y_1, \ldots, Y_n\}$. If P is a reduct of C with respect to D, then $CON_P(D) \geq CON_C(D)$.

Proof: For any $P \subseteq C$ and $U|D = \{Y_1, \ldots, Y_n\}$. If is a reduct of C with respect to D, then it is known from definition (2) that $CON_P(D) \geq CON_C(D)$ for all $Y_j \in U|D$, that is $\sum_{j=1}^{n} CON_P(j) \geq \sum_{j=1}^{n} CON_C(j)$. Hence, $CON_P(D) \geq CON_C(D)$.

Proposition 4: provide an approach to manipulate the monotonicity of the consistent measure under reduction in an incomplete decision system. It should be remark that IDS is consistent if the generalization function $\delta_C(u) = \{\rho(v, d): d \in D, \text{ for some } v \in S_{P,\alpha}(u)\}$ has $|\delta_C(u)| = 1$. i.e. $S_{P,\alpha}(u) \cap [d] = S_{P,\alpha}(u)$ therefor:

$$\mu_P(u_i) = \left(\upsilon\left(S_{P,\alpha}(u_i), Y_1\right), \upsilon\left(S_{P,\alpha}(u_i), Y_2\right), \ldots, \upsilon\left(S_{P,\alpha}(u_i), Y_n\right)\right) =$$

$$(0, 0, \ldots, 1, 0, \ldots, 0)\upsilon\left(S_{P,\alpha}(u_i), Y_j\right) = 1 \text{iff} Y_j = [d]$$

The following algorithms shows the consistent degree tolerance rough set algorithm that implements the idea presented above, algorithm 3 illustrate the consistent degree algorithm:

Algorithm 3; Feature reduction algorithm based consistent degree:

Input C, the set of all condition features, D the set of decision features with n classes

Output: R the reduced feature subset

1. Construct SC, $\alpha(u_i)\forall u_i \in U$[Lin Sun]
2. $R \leftarrow$ null, $T \leftarrow$ null
3. $Con \leftarrow 0$, $\gamma_{best} \leftarrow 0$, $\gamma_T(D) \leftarrow 0$, $Con_T \leftarrow 0$
4. While $\gamma_C(D) \neq \gamma_T(D)$
5. $T \leftarrow R$
6. $\gamma_T(D) \leftarrow \gamma_{best}$
7. If $((C\text{-}R \neq \varnothing)$and $a \in (C\text{-}R)$ do
   a. $Sum_j \leftarrow 0$
   b. For$j \leftarrow 0$, $j<n$, do
      i. Calculate $CON_{R\cup\{a\}}(j)$
      ii. $sum_j \leftarrow sum_j + CON_{R\cup\{a\}}(j)$
   c. If $sum_j \geq Con_T(D)$, do
      i. $T \leftarrow R\cup\{a\}$
      ii. $Con_T \leftarrow sum_j$
      iii. $\gamma_{best} \leftarrow \gamma_T(D)$
      iv. $Con \leftarrow +sum_j$
   d. $R \leftarrow T$
8. Return R

The algorithm services the consistency degree and the dependency value. The dispensable features are added to the current reduct elements. The consistency degree is only used to select the features subset. The termination case is the tolerance rough set dependency value the termination happen when the addition of any remaining feature does not result in an enhance in the dependency.

RESULTS AND DISCUSSION

Expermint1 (thyroid diagnosis): The medical database of thyroid records are grouped in the UCI repository (Bache and Lichman, 2013) collected by Garvin Institute of Medical Research. There are 30 different features that are applied for each patient. Unfortunately, this dataset contains some unknown attribute values. The diagnosis of hypothyroidism represent the main target of this experiment. The dataset depends on recognizing the thyroid hormone quantity in the blood. Hence, the normal ranges for thyroid hormones should be considered and recognized. The classes of thyroid diagnosing hyperthyroid, primary hypothyroid, compensated hypothyroid, secondary hypothyroid and negative hypothyroidism are considered in the classification task. Also, the dataset trust on the blood test of (T4, T3, TSH, TT4, TBG, T4U, FTI) which contains some missing and don't care values.

The efficiency of method of TRSM+ Consistency measure in feature selection is tested wherever the local rules of diagnosis pattern is recognized. Also a comparative study among the TRSM+Consistency measure, TRSM (Radwan and Assiri, 2013) and the machine learning algorithm C4.5 (Ibrahim *et al*., 2011) should be tested. In this study, the proposed method is run on about 2514 (1667 training-838 test) records of relevant data about the patient. By applying Alg1 and Alg2, the decision attribute is dispensable one, Consistent data set, as depicted in Fig. 2.

By applying the TRSM reduction algorithm, an automated rule-based tolerance rough set system in diagnosis creates about nine IF-THEN rules. Statistics are summarizing that accuracy is 97.49%. By Applying TRSM with+Consistency measure, the accuracy of automated rule-based system is increases to be around 98.1 %. The algorithm is running on a training data about 500 data sets of patients. The algorithm select the two high fitness parents (Rules). Common part between two parents called head; The head of two rules is TSH>6 as.

For each data set and algorithms results indicate that (Consistency Measure+TRSM) decreasing size of tree, rule extracted and time taken to build model. Hence, the overall accuracy is increased. Evaluating the accuracy of this model is applying by running inductive machine learning algorithm and genetic algorithm on 500 different cases of hypothyroid patient's as shown in Fig. 3.

The descriptions of the rules generated trsm+ consistency measure:

- Semantic meaning: If (TSH> 6 and FTI< = 64 and TSH = Normal and T4U = Normal and thyroid surgery = FALSE)→Primary Hypothroid

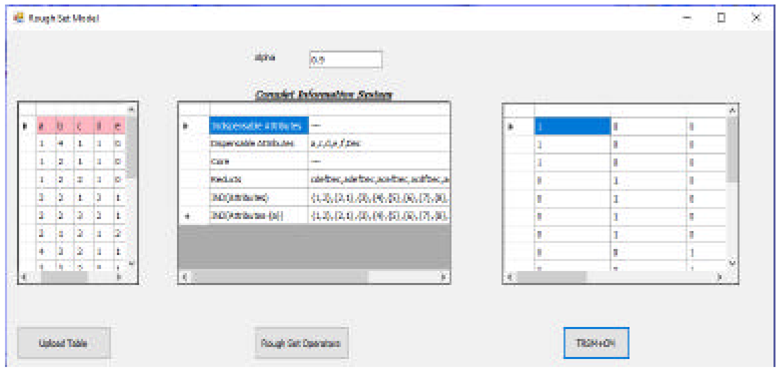

Fig. 2: The algorithm for computing the minimal description set, maximal description set and superfluous attributes

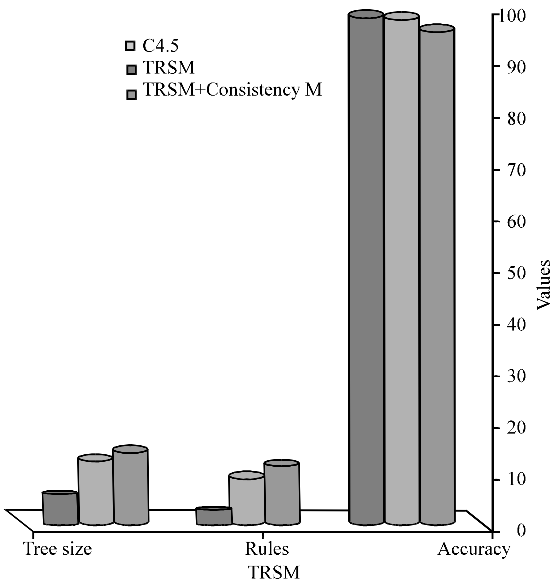

Fig. 3: A comparative study between TRSM and consistency measure

- Rule 2 (Parent 2): If (TSH >6 and FTI< = 64 and on-throxine = FALSE and TSH = Normal and thyroid-surgery = FALSE and TT4 < = 150 and TT4 = Normal and TSH < = 47)→Compensated
- Head TSH > 6
- Child 1: If (TSH >6 and FTI < = 64 and on-throxine = FALSE and T4U = Normal and surgery = FALSE)→Compensated
- Child 2: If (TSH >6 and FTI < = 64 = and TSH = Normal and thyroid-surgery = FALSE and TT4< = 150 and TT4 = Normal and TSH < = 47)→Primary

Expermint 2 (breast cancer): The breast cancer dataset, gathered from the UCI machine learning repository (Bache and Lichman, 2013) is tested. It is constructed from 9 (clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitoses) features and classification class (2 for benign, 4 for malignant). The dataset (738 records with 16 missing value) are divided in two equal parts (one for the training data and one for the test data) (Fig. 4).

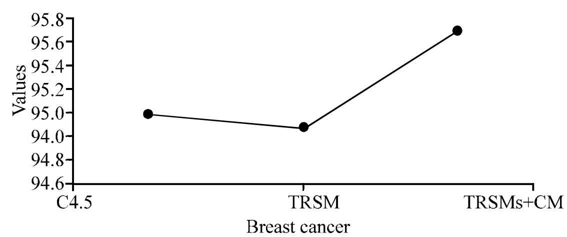

Fig. 4: A classification accuracy comparison among C4.5, TRSM and consistency measure

By applying the TRSM+Consistency measure, The breast cancer rule depends on TRSM+CM has two core attributes (the reduct is uniformity size and epithelial size).

The descriptions of the rules generated TRSM+ consistency measure for breast cancer:

- Rule 1: Uniformity-size = normal and epithelial-size = upnormal→ benign
- Rule 2: Uniformity-size = normal and epithelial-size = upnormal→ malignant
- Rule 3: Uniformity-size = normal and epithelial-size = upnormal→ malignant
- Rule 4: Uniformity-size = normal and epithelial-size = upnormal→ benign
- Rule 5: Uniformity-size = normal and epithelial-size = upnormal→ benign

In contrast with other different methods such as tolerance rough set and the machine learning algorithm C4.5, TRSM+Consistency measure produce high classification than others as illustrated.

CONCLUSION

Knowledge acquisition is considered as a rich area of research where the experts experiment should be saved. Unfortunately most dataset are appeared in the learning process irrelevant, high dimensionality, uncertain and contain missing attribute value. Although, many algorithms achieve high performance in increasing accuracy they still depend on the case study. Thus a hybrid model that achieve robust classification is needed. By this study, tolerance rough sets with consistent measure are integrated with step-deck coordination function in purpose of reducing the computation time, solving of the problem of imbalance classes and achieving high classification accuracy. By this study, tolerance rough sets with consistent measure is studied to solve the problems of missing attribute values and imbalance classes. Moreover, the optimization problem based on the consistency distribution is constructed where the variation is considered as the optimization function. This study is applied on the medical dataset (breast cancer and thyroid datasets) to declare the results. The comparison study illustrates the ability of the new model. In future research, this model should be accompanied with the theory of credibility to improve the classification accuracy.

REFERENCES

Ali, R., M.H. Siddiqi and S. Lee, 2015. Rough set-based approaches for discretization: A compact review. Artif. Intell. Rev., 44: 235-263.

Azar, A.T., H.H. Inbarani and K.R. Devi, 2017. Improved dominance rough set-based classification system. Neural Comput. Appl., 28: 2231-2246.

Bache, K. and M. Lichman, 2013. UCI machine learning repository. University of California, School of Information and Computer Science, Irvine, CA.http://archive.ics.uci.edu/ml/citation_policy.html.

Chen, D., D.W. Cui, C.X. Wang and Z.R. Wang, 2006. A rough set-based hierarchical clustering algorithm for categorical data. Intl. J. Inf. Technol., 12: 149-159.

Dai, J. and H. Tian, 2013. Entropy measures and granularity measures for set-valued information systems. Inf. Sci., 240: 72-8

Dai, J., D. Huang, H. Su, H. Tian and T. Yang, 2014. Uncertainty measurement for covering rough sets. Intl. J. Uncertainty, Fuzziness Knowledge Based Syst., 22: 217-233.

Herawan, T., I.T.R. Yanto and M.M. Deris, 2010. A Construction of Hierarchical Rough Set Approximations in Information Systems using Dependency of Attributes. In: Advances in Intelligent Information and Database Systems: Studies in Computational Intelligence, Nguyen, N.T., R. Katarzyniak and S.M. Chen (Eds.). Springer, Berlin, Germany, ISBN:978-3-642-12089-3, pp: 3-15.

Ho, T.B. and N.B. Nguyen, 2002. Nonhierarchical document clustering based on a tolerance rough set model. Int. J. Intell. Syst., 17: 199-212.

Hu, Y.C., 2015. Flow-based tolerance rough sets for pattern classification. Appl. Soft Comput., 27: 322-331.

Huang, Y., T. Li, C. Luo, H. Fujita and S.J. Horng, 2017. Dynamic variable precision rough set approach for probabilistic set-valued information systems. Knowl. Based Syst., 122: 131-147.

Ibrahim, N., T. Hamza and E. Radwan, 2011. An evolutionary machine learning algorithm for classifying thyroid diseases diagnoses. Egyptian Comput. Sci. J., 35: 73-86.

Jusoh, N.A., J.M. Zain, N.K.N. Ismail and T.A.A. Kadir, 2007. Comparison between techniques in feature extraction. Proceedings of the International Conference on Electrical Engineering and Informatics, June 17-19, 2007, Institut Teknologi Bandung, Bandung, Indonesia, ISBN:978-979-16338-0-2, pp: 973-975.

Mac Parthalain, N. and Q. Shen, 2009. Exploring the boundary region of tolerance rough sets for feature selection. Pattern Recognition, 42: 655-667.

Radwan, E. and A.M. Assiri, 2013. Thyroid diagnosis based technique on rough sets with modified similarity relation. Intl. J. Adv. Comput. Sci. Appl., 4: 120-126.

Ramanna, S., 2011. Near tolerance rough sets. Appl. Math. Sci., 5: 1895-1902.

Rivero, L.C., J.H. Doorn and V.E. Ferraggine, 2005. Rough Sets: Encyclopedia of Database Technologies and Applications. Idea Group Reference, Hershey, London, ISBN:9781591407959, Pages: 784.

Sun, L., J. Xu and Y. Li, 2014. A feature selection approach of inconsistent decision systems in rough set. J. Comput., 9: 1333-1340.

Sun, L., J. Xu and Y. Tian, 2012. Feature selection using rough entropy-based uncertainty measures in incomplete decision systems. Knowl. Based Syst., 36: 206-216.

Wang, F., J. Liang and C. Dang, 2013. Attribute reduction for dynamic data sets. Appl. Soft Comput., 13: 676-689.

Wu, C. and L. Wang, 2008. An improved limited and variable precision rough set model for rule acquisition based on pansystems methodology. Kybernetes, 37: 1264-1271.

Wu, J., D.Z. Tian, L. Wang and S.J. Yan, 2009. Rough set model based on credibility measure. Proceedings of the 2009 International Conference on Machine Learning and Cybernetics Vol. 3, July 12-15, 2009, IEEE, Hebei, China, ISBN:978-1-4244-3702-3, pp: 1822-1827.

Wu, W.Z., Y. Qian, T.J. Li and S.M. Gu, 2017. On rule acquisition in incomplete multi-scale decision tables. Inf. Sci., 378: 282-302.