Chapter 4

Feature subset selection of semi-supervised data: An Intuitionistic Fuzzy-Rough set based concept

4.1 Introduction

Due to generation of huge amount of data from various information technology sources, many of the decision class labels are found to be missing in the high-dimensional data sets such as gene expression microarray data [91, 4]. On the basis of decision class labels, feature selection techniques can be divided into three categories viz. supervised, semi-supervised and unsupervised [44, 57, 95, 107]. Supervised feature selection techniques [22] are applied to the data sets with all the decision class labels available and unsupervised feature selection concepts [23, 64] are practised to the data sets with no decision class labels while semi-supervised attribute reduction methods are applied to the data sets with combination of available decision class labels and missing decision class labels. Supervised learning approaches learn underlying functional relationship available in data, while unsupervised learning concepts use some inherent structure available in data and find some groups in the data such that objects in the same group are different from the objects of the other group on the basis of some criteria. Traditional learning methods are unable to exploit unlabelled data for pattern recognition and knowledge discovery. Therefore, semi-supervised learning approaches can play vital role in order to deal with the information system containing both labelled and unlabelled data.

Fuzzy rough set theory has been successfully applied to deal with real-valued data sets in order to reduce the dimension of the data set, especially for feature selection. Very few researches have been presented semi-supervised feature selection based on fuzzy rough set theory. In the recent years, some of the intuitionistic fuzzy rough set models [8, 13, 105, 109, 110] were proposed and successfully implemented for pattern recognition and decision making [34, 70, 93, 94, 98, 101, , 102, 103, 106]. In this study, we present a novel intuitionistic fuzzy rough set assisted feature selection which can easily deal with the data set having both labelled and unlabelled data. Furthermore, we propose theorems supporting our concept and prove the validity of the theorems. Moreover, we propose an algorithm based on our proposed method. Finally, we apply this approach to an information system containing both labelled and unlabelled data and show that it performs better than semi-supervised fuzzy rough feature selection approach proposed by Jensen et.al. [44].

4.2 Intuitionistic fuzzy-rough feature selection

Concept of feature selection based on intuitionistic fuzzy rough set can be extended as follows: A subset B of set of conditional attributes C can be defined using intuitionistic fuzzy similarity relation as follows:

$$\langle \mu_{R_B(x,y)}, \nu_{R_B(x,y)} \rangle = T(\langle \mu_{R_a}(x,y), \nu_{R_a}(x,y) \rangle), \forall a \in B \tag{4.1}$$

where, $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are two intuitionistic fuzzy values and T is an intuitionistic fuzzy triangular norm or t-norm. Now, lower and upper approximations of an intuitionistic fuzzy set A in X (universe of discourse) based on intuitionistic fuzzy similarity relation R is defined as follows [13]:

$$(R_B \downarrow_I A(x)) = inf_{y \in X} I(R_B(x,y), A(y)), \forall x, y \in X \tag{4.2}$$

$$(R_B \uparrow_T A(x)) = sup_{y \in X} T(R_B(x,y), A(y)), \forall x, y \in X \tag{4.3}$$

where, T and I are intuitionistic fuzzy triangular norm and intuitionistic fuzzy implicator respectively. Now, on the basis of above defined lower approximation, we can define intuitionistic fuzzy positive region by:

$$pos_B(x) = (R_B \downarrow_I [x]_d)(x) \tag{4.4}$$

where, $[x]_d$ contains all objects having same decision value as x . Now, we consider following intuitionistic fuzzy triangular norm T_w and intuitionistic fuzzy implicator I_w as mentioned in [13]:

$$T_w(x,y) = \langle max(0, x_1 + y_1 - 1), min(1, x_2 + y_2) \rangle \tag{4.5}$$

$$I_w(x,y) = \langle min(1, 1 + y_1 - x_1, 1 + x_2 - y_2), max(0, y_2 - x_2) \rangle \tag{4.6}$$

where, $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are two intuitionistic fuzzy values. For every data instance, we can redefine intuitionistic fuzzy positive region as follows

$pos_B(x) = (R_B \downarrow_I [x]_d)(x) = inf_{y \in X} I(R_B(x,y), [x]_d(y))$

$=\min\{inf_{y \in [x]_d} I(R_B(x,y), [x]_d(y)), inf_{y \notin [x]_d} I(R_B(x,y), [x]_d(y))\}$

$=\min\{inf_{y \in [x]_d} I(\langle \mu_{R_B(x,y)}, \nu_{R_B(x,y)} \rangle, \langle 1, 0 \rangle),$

$inf_{y \notin [x]_d} I(\langle \mu_{R_B(x,y)}, \nu_{R_B(x,y)} \rangle, \langle 0, 1 \rangle)\}$

$= \min\ \{inf_{y \in [x]_d} \langle min(1, 1 + 1 - \mu_{R_B}(x,y), 1 + \nu_{R_B}(x,y) - 0), max(0, 0 - \nu_{R_B}(x,y)) \rangle,$

$inf_{y \notin [x]_d} \langle min(1, 1 + 0 - \mu_{R_B}(x,y), 1 + \nu_{R_B}(x,y) - 1), max(0, 1 - \nu_{R_B}(x,y)) \rangle\}$

$= \min\ \{inf_{y \in [x]_d} \langle 1, 0 \rangle, inf_{y \notin [x]_d} \langle \nu_{R_B}(x,y), 1 - \nu_{R_B}(x,y) \rangle\}$

$= \min\ \{\langle 1, 0\rangle, inf_{y \notin [x]_d} \langle \nu_{R_B}(x, y), 1 - \nu_{R_B}(x, y)\rangle\}$

$$pos_B(x) = inf_{y \notin [x]_d} \langle \nu_{R_B}(x, y), 1 - \nu_{R_B}(x, y)\rangle \tag{4.7}$$

Let $IFDS = \{X, C \cup q, V_{IF}, IF\}$ be an intuitionistic fuzzy decision system. So, we define degree of dependency of decision feature on subset of conditional features as follows:

$$\Gamma_B = \frac{|pos_B|}{|X|}$$

where, $|.|$ in numerator is the cardinality of an intuitionistic fuzzy set as defined in chapter 1 and in denominator, it denotes cardinality of a crisp set.

4.3 Semi-supervised Intuitionistic Fuzzy-Rough Feature Selection

It is very expensive and time consuming for data experts to deal large number of labelled data, this motivates us for some better technique viz. semi-supervised techniques in order to learn about small amounts of labelled data and larger amounts of unlabelled data. For handling both labelled and unlabelled data, some modifications are required in the definition of positive region can be given as follows:

Theorem 4.3.1 *Let L and U be the sets of labelled and unlabelled objects respectively and $\{L, U\}$ is a partition of X (universe of discourse), i.e. $L \cap U = \phi$ and $L \cup U = X$, then positive region in the system can be defined by:*

$$pos_B^{ssl}(x) = \begin{cases} inf_{y \neq x} \langle \nu_{R_B}(x, y), 1 - \nu_{R_B}(x, y)\rangle, & if\ x \in U \\ inf_{y \in (U \cup co[x]_d^L)} \langle \nu_{R_B}(x, y), 1 - \nu_{R_B}(x, y)\rangle, & if\ x \in L \end{cases} \tag{4.8}$$

where, $[x]_d^L$ represents the set of labelled objects having same decision value as x and $co(.)$ is the complement operator.

Proof:

Let $x \in U$, then its decision class contains only x . Now Eq. (4.7) instantaneously simplifies to

$$pos_B^{ssl}(x) = inf_{y \neq x} \langle \nu_{R_B}(x,y), 1-\nu_{R_B}(x,y) \rangle \ .$$

If $x \in L$, then decision class of x consists of all labelled instances y satisfying $d(x) = d(y)$. All unlabelled objects are not element of it, as all of them belong to their own individual classes. Therefore, infimum is taken over $U \cup co[x]_d^L$ and it results in $pos_B^{ssl}(x) = inf_{y \in (U \cup co[x]_d^L)} \langle \nu_{R_B}(x,y), 1-\nu_{R_B}(x,y) \rangle$. Hence we get the required result. Now, new degree of dependency can be defined as:

$$\Gamma_B^{ssl} = \frac{|pos_B^{ssl}|}{|X|}$$

Theorem 4.3.2 *For every $B \subseteq C, \Gamma_B^{ssl} \leq \Gamma_B$.*

Proof

For any given function f and sets R and S along with condition $R \subseteq S$, it is obvious that

$$inf_{x \in S} f(x) \leq inf_{x \in R} f(x) \tag{4.9}$$

If $x \in X$, then, for any semi-supervised model, either $x \in U$ or $x \in L$. Let $x \in U$, then according to Eq. (4.8), we can conclude that:

$pos_B^{ssl}(x) = inf_{y \neq x} \langle \nu_{R_B}(x,y), 1-\nu_{R_B}(x,y) \rangle$

$= inf_{y \in (X \setminus \{x\})} \langle \nu_{R_B}(x,y), 1-\nu_{R_B}(x,y) \rangle$

$\leq inf_{y \in co([x]_d)} \langle \nu_{R_B}(x,y), 1-\nu_{R_B}(x,y) \rangle$ *(using Eq. (4.9) along with $co([x]_d) \subseteq (X \setminus \{x\})$)*

$= inf_{y \notin [x]_d} \langle \nu_{R_B}(x,y), 1-\nu_{R_B}(x,y) \rangle$

where $[x]_d$ is established within the completely labelled system. Now, let $x \in L$, then

$pos_B^{ssl}(x) = inf_{y \in (U \cup co[x]_d^L)} \langle \nu_{R_B}(x,y), 1-\nu_{R_B}(x,y) \rangle$

$\leq inf_{y \in co[x]_d} \langle \nu_{R_B}(x,y), 1-\nu_{R_B}(x,y) \rangle$ (using Eq. (4.9) along with $co([x]_d) \subseteq (U \cup co[x]_d^L)$)

$= inf_{y \notin [x]_d} \langle \nu_{R_B}(x,y), 1-\nu_{R_B}(x,y) \rangle = pos_{(x)}$

Now, in semi-supervised model either objects are labelled or have no label. So, we can conclude that

$\forall x \in X, pos_B^{ssl}(x) \leq pos_(x)$.

Therefore, $\dfrac{|pos_B^{ssl}(x)|}{|X|} \leq \dfrac{|pos_B(x)|}{|X|}$. Hence, $\Gamma_B^{ssl} \leq \Gamma_B$.

4.4 Algorithm for Semi-supervised intuitionistic fuzzy rough feature selection

In this section, we give the algorithm for feature selection using semi-supervised intuitionistic fuzzy rough set technique as follows:

Input : C, Collection of all conditional attributes;

Output : Z, the reduct set;

$Z \leftarrow \{\}; \Gamma_{best}^{ssl} = 0;$

do

$L \leftarrow Z$

$\forall p \in (C \backslash B)$

if $\Gamma_{Z \cup \{p\}}^{ssl} \geq \Gamma_L^{ssl}$

$L \leftarrow Z \cup \{p\}$

$\Gamma_{best}^{ssl} = \Gamma_L^{ssl}$

$Z \leftarrow L$

while $\Gamma_{best}^{ssl} \neq \Gamma_C^{ssl}$ return Z

4.5 Worked Example

An arbitrary example of fuzzy information system inspired from Jensen et.al. [40] is given in Table 4.1 with universe of discourse $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, set of conditional attributes $C = \{a, b, c, d, e, f\}$ and one decision attribute $\{q\}$.

Table 4.1: Semi-Supervised Fuzzy Information System

Objects \ Attributes	a	b	c	d	e	f	q
x_1	0.4	0.4	1.0	0.8	0.4	0.2	-
x_2	0.6	1.0	0.6	0.8	0.2	1.0	0
x_3	0.8	0.4	0.4	0.6	1.0	0.2	1
x_4	1.0	0.6	0.2	1.0	0.6	0.4	0
x_5	0.2	1.0	0.8	0.4	0.4	0.6	-
x_6	0.6	0.6	0.8	0.2	0.8	0.8	1

Now, similarity degree of two objects can be calculated using following similarity relation [35] :

$$R_a(x, y) = 1 - \frac{|\mu_a(x) - \mu_a(y)|}{|\mu_{a_{max}} - \mu_{a_{min}}|} \tag{4.10}$$

where, $\mu_a(x), \mu_a(y)$ are membership grades of objects x, y respectively and $\mu_{a_{max}}, \mu_{a_{min}}$ are maximum and minimum membership grades for an attribute a respectively.

Now, we can calculate degree of dependency of decision feature q over conditional feature $\{a\}$ using [42] as follows:

$\gamma^{ssl}_{\{a\}} = \frac{0.25+0.25+0.25+0.25+0.25+0}{6} = \frac{1}{6}$

Similarly, dependency functions over $\{b\}, \{c\}, \{d\}, \{e\}$ and $\{f\}$ are :

$\gamma^{ssl}_{\{b\}} = \frac{0+0+0+0.33+0+0.33}{6} = \frac{0.66}{6}$

$\gamma^{ssl}_{\{c\}} = \frac{0.25+0.25+0.25+0.25+0+0}{6} = \frac{1}{6}$

$\gamma^{ssl}_{\{d\}} = \frac{0+0+0.25+0.25+0.25+0.25}{6} = \frac{1}{6}$

$\gamma^{ssl}_{\{e\}} = \frac{0+0.25+0.25+0.25+0.25+0.25}{6} = \frac{1.25}{6}$

$\gamma^{ssl}_{\{f\}} = \frac{0+0+0.50+0.25+0.25+0.25}{6} = \frac{1.25}{6}$

Since $\{e\}$ and $\{f\}$ have same and the highest degree of dependency values, so we can take any one of them as reduct candidate. Taking $\{e\}$ as reduct candidate, we add other attributes one by one and find degree of dependencies as follows:

$\gamma^{ssl}_{\{a,e\}} = \frac{2.25}{6}, \gamma^{ssl}_{\{b,e\}} = \frac{2.33}{6}, \gamma^{ssl}_{\{c,e\}} = \frac{2.50}{6}, \gamma^{ssl}_{\{d,e\}} = \frac{2.25}{6}, \gamma^{ssl}_{\{e,f\}} = \frac{2.25}{6}$.

Now, we insert other attributes to the next reduct candidate i.e. $\{c, e\}$ and get degree of dependencies as:

$\gamma^{ssl}_{\{a,c,e\}} = \frac{2.50}{6}, \gamma^{ssl}_{\{b,c,e\}} = \frac{2.84}{6},$

$\gamma^{ssl}_{\{c,d,e\}} = \frac{3.00}{6}, \gamma^{ssl}_{\{c,e,f\}} = \frac{3.50}{6}.$

Since $\{c, e, f\}$ provides maximum value of degree of dependency. Hence, other attributes are added to the potential reduct set c, e, f and corresponding degree of dependencies are:

$\gamma^{ssl}_{\{a,c,e,f\}} = \frac{3.50}{6}, \gamma^{ssl}_{\{b,c,e,f\}} = \frac{3.92}{6}, \gamma^{ssl}_{\{c,d,e,f\}} = \frac{3.50}{6}$

So, we get $\{b, c, e, f\}$ as next potential reduct set and after adding rest of the attributes to this set, we obtain degree of dependencies as follows:

$\gamma^{ssl}_{\{a,b,c,e,f\}} = \frac{3.92}{6}, \gamma^{ssl}_{\{b,c,d,e,f\}} = \frac{3.92}{6}$

On adding other attributes to the potential reduct set $\{b, c, e, f\}$, we get no increment in degree of dependency. Hence, the final reduct is $\{b, c, e, f\}$.

Now we convert the above fuzzy information system into intuitionistic fuzzy information system by using Jurio et. al. [46] concept with fixed hesitancy degree as 0.2. The reduced information system is given in Table 4.2.

Table 4.2: Semi-Supervised Intuitionistic Fuzzy Information System

Objects \ Attributes	a	b	c	d	e	f	q
x_1	$\langle 0.32, 0.48 \rangle$	$\langle 0.32, 0.48 \rangle$	$\langle 0.80, 0.00 \rangle$	$\langle 0.64, 0.16 \rangle$	$\langle 0.32, 0.48 \rangle$	$\langle 0.16, 0.64 \rangle$	-
x_2	$\langle 0.48, 0.32 \rangle$	$\langle 0.80, 0.00 \rangle$	$\langle 0.48, 0.32 \rangle$	$\langle 0.64, 0.16 \rangle$	$\langle 0.16, 0.64 \rangle$	$\langle 0.80, 0.00 \rangle$	0
x_3	$\langle 0.64, 0.16 \rangle$	$\langle 0.32, 0.48 \rangle$	$\langle 0.32, 0.48 \rangle$	$\langle 0.48, 0.32 \rangle$	$\langle 0.80, 0.00 \rangle$	$\langle 0.16, 0.64 \rangle$	1
x_4	$\langle 0.80, 0.00 \rangle$	$\langle 0.48, 0.32 \rangle$	$\langle 0.16, 0.64 \rangle$	$\langle 0.80, 0.00 \rangle$	$\langle 0.48, 0.32 \rangle$	$\langle 0.32, 0.48 \rangle$	0
x_5	$\langle 0.16, 0.64 \rangle$	$\langle 0.80, 0.00 \rangle$	$\langle 0.64, 0.16 \rangle$	$\langle 0.32, 0.48 \rangle$	$\langle 0.32, 0.48 \rangle$	$\langle 0.48, 0.32 \rangle$	-
x_6	$\langle 0.48, 0.32 \rangle$	$\langle 0.48, 0.32 \rangle$	$\langle 0.64, 0.16 \rangle$	$\langle 0.16, 0.64 \rangle$	$\langle 0.64, 0.16 \rangle$	$\langle 0.64, 0.16 \rangle$	1

Above defined fuzzy similarity degree gives an idea for intuitionistic fuzzy similarity degree. Now, We can define an intuitionistic fuzzy tolerance relation using [35] as follows:

Let $\alpha = 1 - \frac{|\mu_a(x)-\mu_a(y)|}{|\mu_{a_{max}}-\mu_{a_{min}}|}, \beta = \frac{|\nu_a(x)-\nu_a(y)|}{|\nu_{a_{max}}-\nu_{a_{min}}|},$

where $\mu_a(x)$,$\mu_a(y)$ and $\nu_a(x)$, $\nu_a(y)$ are the membership and non-membership grades of x and y in U for any attribute $a \in P$ and $\mu_{a_{max}}$, $\mu_{a_{min}}$ and $\nu_{a_{max}}$, $\nu_{a_{min}}$ are the maximum and minimum membership and non-membership grades respectively that attribute a may take. Then,

$$\langle \mu_{R_a}(x,y), \nu_{R_a}(x,y)\rangle = \begin{cases} \langle \alpha, \beta\rangle, & if\ \alpha+\beta \leq 1 \\ \langle 1, 0\rangle, & if\ \alpha+\beta > 1 \end{cases} \tag{4.11}$$

where, μ_{R_a} and ν_{R_a} are membership and non-membership grades of intuitionistic fuzzy tolerance relation.

If R_P is the intuitionistic fuzzy tolerance relation induced by the subset of features P, then,

$$\langle \mu_{R_P}(x,y), \nu_{R_P}(x,y)\rangle = inf_{a\in P}\langle \mu_{R_a}(x,y), \nu_{R_a}(x,y)\rangle \tag{4.12}$$

Now, we calculate the reduct set of intuitionistic fuzzy information system as given in Table 4.2 by using above section as follows:

$\nu_{R_P}(x,y)$ can be calculated by using Eq. (4.12) and recorded in Table 4.3.

Table 4.3: Semi-Supervised Intuitionistic Fuzzy Relation

Objects	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0	0.25	0.50	0.75	0.25	0.25
x_2	0.25	0	0.25	0.50	0.50	0
x_3	0.50	0.25	0	0.25	0.75	0.25
x_4	0.75	0.50	0.25	0	1	0.50
x_5	0.25	0.50	0.75	1	0	0.50
x_6	0.25	0	0.25	0.50	0.50	0

Now, positive region for object x_1 over attribute $\{a\}$ can be given as:

$pos^{ssl}_{\{a\}}(x_1) = inf(\langle 0.25, 0.75\rangle, \langle 0.50, 0.50\rangle, \langle 0.75, 0.25\rangle, \langle 0.25, 0.75\rangle, \langle 0.25, 0.75\rangle)$

$= \langle 0.25, 0.75 \rangle$

Similarly positive region for other objects can be given by:

$pos^{ssl}_{\{a\}}(x_2) = \langle 0, 1 \rangle, pos^{ssl}_{\{a\}}(x_3) = \langle 0.25, 0.75 \rangle, pos^{ssl}_{\{a\}}(x_4) = \langle 0.25, 0.75 \rangle,$

$pos^{ssl}_{\{a\}}(x_5) = \langle 0.25, 0.75 \rangle, pos^{ssl}_{\{a\}}(x_2) = \langle 0, 1 \rangle$

Now, we can calculate degree of dependency of decision feature q over conditional feature $\{a\}$ using [34] as follows:

$\gamma^{ssl}_{\{a\}} = \frac{0.25+0+0.25+0.25+0.25+0}{6} = \frac{1}{6}$

Similarly, dependency functions over $\{b\}, \{c\}, \{d\}, \{e\}$ and $\{f\}$ are :

$\gamma^{ssl}_{\{b\}} = \frac{0+0+0+0.33+0+0.33}{6} = \frac{0.66}{6}$

$\gamma^{ssl}_{\{c\}} = \frac{0.25+0.25+0.25+0.25+0+0}{6} = \frac{1}{6}$

$\gamma^{ssl}_{\{d\}} = \frac{0+0+0.25+0.25+0.25+0.25}{6} = \frac{1}{6}$

$\gamma^{ssl}_{\{e\}} = \frac{0+0.25+0.25+0.25+0.25+0.25}{6} = \frac{1.25}{6}$

$\gamma^{ssl}_{\{f\}} = \frac{0+0+0.50+0.25+0.25+0.25}{6} = \frac{1.25}{6}$

Since $\{e\}$ and $\{f\}$ have same and the highest degree of dependency values, so we can take any one of them as reduct candidate. Taking $\{e\}$ as reduct candidate, we add other attributes one by one and find degree of dependencies as follows:

$\gamma^{ssl}_{\{a,e\}} = \frac{2.25}{6}, \gamma^{ssl}_{\{b,e\}} = \frac{2.33}{6}, \gamma^{ssl}_{\{c,e\}} = \frac{2.25}{6}, \gamma^{ssl}_{\{d,e\}} = \frac{2.25}{6}, \gamma^{ssl}_{\{e,f\}} = \frac{2.25}{6}$ Now, we insert other attributes to the next reduct candidate i.e. $\{b, e\}$ and get degree of dependencies as:

$\gamma^{ssl}_{\{a,b,e\}} = \frac{3.17}{6}, \gamma^{ssl}_{\{b,c,e\}} = \frac{2.83}{6},$

$\gamma^{ssl}_{\{b,d,e\}} = \frac{3.68}{6}, \gamma^{ssl}_{\{b,e,f\}} = \frac{3.08}{6}$

Since $\{c, e, f\}$ provides maximum value of degree of dependency. Hence, other attributes are added to the potential reduct set $\{b, d, e\}$ and corresponding degree of dependencies are:

$\gamma^{ssl}_{\{a,b,d,e\}} = \frac{3.17}{6}, \gamma^{ssl}_{\{b,c,d,e\}} = \frac{3.08}{6},$

$\gamma^{ssl}_{\{b,d,e,f\}} = \frac{3.68}{6}$

Now, we get no increment in degree of dependency. So, process exits and we obtain the reduct set as $\{b, d, e\}$.

4.6 Conclusion

Semi-supervised approaches are essential to deal with abundance of unlabelled data available in high-dimensional data set as it is often costly and heavily time consuming for domain experts to find decision class labels. This study has proposed a novel concept to feature selection for data set with labelled and unlabelled data. The proposed approach provides a valid reduct even if the maximum number of the data class labels are missing. In this paper, we presented an intuitionistic fuzzy rough set model and generalized it for attribute selection for semi-supervised data. Furthermore, we proposed supporting theorems and proved their validity. Moreover, an algorithm has been presented in order to demonstrate our approach. Finally, the proposed algorithm applied to an example data set. We observed that our proposed approach is performing better than previously reported semi-supervised fuzzy rough feature selection in terms of selected attributes.

****