Chapter 2

Background

A dataset is the collection of objects or instances or samples, each of which is characterised by set of properties. These set of properties are referred to as features. This collection of all the objects is called universe of discourse and is denoted by U. Let $U = \{x_1, x_2, \ldots x_n\}$ be non empty collection of objects $x_i \in U$. Similarly, let C be non empty set of features such that for any $a \in C, a : U \implies V_a$, where V_a is the set of all possible values taken by feature a. These two sets cummulatively i.e. pair (U, C) forms an information system.

2.1 Rough Set Theory

Rough set theory as proposed by Pawlak [176] is an extension of classical set theory that can be utilized to mine knowledge from data. It is based on the concept of indiscernibility. It can be effectively applied to reduce the size of data while retaining the actual information content. Suppose (U, C) is an information system, then an equivalence relation R_A is associated with each feature subset $A \in C$ given by:

$$R_A = \{(x, y) \in U \times U \mid \forall a \in A, a(x) = a(y)\} \tag{2.1}$$

where $a(x)$ is the value of feature a for an instance or object x. Two objects x and y are indistinguishable or indiscernible with respect to feature subset A if $(x, y) \in R_A$. The corresponding equivalence class $[x]_A$ consists of all the objects related to x according to R_A. So, any set of objects $X \subseteq U$ can be approximated by a pair of concepts employing the above defined equivalence relation as:

$$R \downarrow_A X = \{x \mid [x]_A \subseteq X\} \tag{2.2}$$

$$R \uparrow_A X = \{x \mid [x]_A \cap X \neq \phi\} \tag{2.3}$$

where $R \downarrow_A X$ and $R \uparrow_A X$ denotes the lower and upper approximation respectively. The pair $(R \downarrow_A X, R \uparrow_A X)$ is called the rough set. Lower approximation denotes those objects that belong to X with certainty while objects belonging to upper approximation may possibly be member of X. The difference between these two approximations gives the boundary region.

$$BND_A(X) = R \uparrow_A X - R \downarrow_A X \tag{2.4}$$

A unique form of information system in which one of the feature is class or decision or target label of the instance is called decision system. Let the decision feature be denoted by D. Then, the decision system is $(U, C \cup D)$, where C is the set of conditional features. Union of lower approximations over all the equivalence classes of target label gives the positive region and is defined as:

$$Pos_A(D) = \cup_{X \in U \setminus D} R \downarrow_A \ X \tag{2.5}$$

Employing the value of positive region, the degree of dependence of decision class on feature subset is calculated as:

$$\gamma_A(D) = \frac{\mid Pos_A(D) \mid}{\mid U \mid} \tag{2.6}$$

An informtion system is consistent if $\gamma_A(D) = 1$. This measure is utilized to reckon the quality of feature subset. A feature subset $A \subseteq C$ is reduct if its dependency is same as that of dataset containing entire set of conditional features, i.e. $\gamma_A(D) =$

$\gamma_C(D)$. If this condition is relaxed, then subset A may not be necessarily optimal and is referred to as super reduct.

Feature selection can be effectively done using rough set based approach without requiring any additional information. However, it can only be applied to discrete valued dataset. Thus, real valued data needs to be discretized for applying rough set based approaches which may result in loss of valuable information.

Fuzzy rough set theory [37, 38], an extension of rough set theory is the solution that can be effectively applied to real valued dataset by assigning the membership degree to each objects in the set thereby eliminating the need for discretization. The hybridization of fuzzy set and rough set can resolve both the uncertainities and vagueness arising in the data along with overcoming the limitations of rough set based approaches.

2.2 Fuzzy Set Theory

Instead of assigning a strignent condition on whether an object belongs to a set or not, fuzzy set theory associates a degree of memebership μ with which a object belongs to a set. So, a fuzzy set A can be written as:

$$A = \{\mu_A(x) \mid x \in U\} \tag{2.7}$$

For example, if we want to ask a whether a person is intelligent or not. So, instead of just saying yes or no, fuzzy set allows to answer with partial degree i.e. quite intelligent, very intelligent, etc.

Some of points/terms related to fuzzy set theory:

- A fuzzy relation $R : U \times U \implies [0, 1]$ is fuzzy similarity relation if it is:

  1. Reflexive: $R(x, x) = 1, \forall x \in U$

2. Symmetric: $R(x, y) = R(y, x), \forall x, y \in U$

- The cardinality of fuzzy set A is given as: $\mid A \mid = \sum_{x \in U} A(x)$

- An increasing function $T : [0, 1] \times [0, 1] \implies [0, 1]$ satisfying $T(1, x) = x, \forall x \in [0, 1]$ is called triangular norm or t norm.
  Some of the common t norms widely employed are: $T_L(x, y) = max\{0, x + y - 1\}$, $forx, y \in [0, 1]$ (Lukasiewicz t norm) and $T_M(x, y) = min\{x, y\}$.

- A mapping increasing in its first and decreasing in its second component defined as $I : [0, 1] \times [0, 1] \implies [0, 1]$ satisfying $I(0, 0) = 1$ and $I(1, x) = x, \forall x \in [0, 1]$ is an fuzzy implicator. Some of the widely used implicators are: $I_L(x, y) = min\{1, 1 - x + y\}, \forall x, y \in [0, 1]$ (Lukasiewicz implicator) and $I_M(x, y) = max\{1 - x, y\}$.

2.3 Fuzzy Rough Set Theory

Rough set theory handles uncertainty arising in the data while fuzzy set theory handles the vagueness. Hybridisation of these two theories has therefore got lot of applications namely information retireval, feature selection, etc.
Given any set $X \subseteq U$ and the fuzzy similarity relation R, the lower and upper approximation of X by R is given as:

$$R \downarrow_A X(x) = inf_{y \in U} I(R(x, y), X(y)) \tag{2.8}$$

$$R \uparrow_A X(x) = sup_{y \in U} T(R(x, y), X(y)) \tag{2.9}$$

where I and T are fuzzy implicator and fuzzy t norm respectively. However, the above defined approximations are prone to noise because of presence of infemum and supremum operators.

2.4 Fuzzy Rough Set based Feature Selection

The applicability of rough set for feature selection was limited to dicrete data. Fuzzy rough set based feature selection was introduced as most real life applications containing real valued datasets. The equivalence relation in rough set can be extended to similarity relation in case of fuzzy rough set. And the corresponding approximations (eq (2.8)) are calculated. Based on the value of lower approximation, positive region is reckoned as:

$$Pos_A(D)(x) = sup_{x \in U \setminus D} R \downarrow_A D(x) \tag{2.10}$$

where D is the decision class. Then, the degree of dependency of decision feature D on feature subset A is defined as:

$$\gamma_A(D) = \frac{\mid Pos_A(D)(x) \mid}{\mid U \mid} \tag{2.11}$$

Using the dependency degree, the quality of feature subset is evaluated.

One of he most common method for finding feature subset is by using forward greedy quick reduct algorithm. It iteratively adds feature producing highest degree of dependency to reduct set until some termination condition is reached. The algorithm for quick reduct is as follows:

Reduct Algorithm (C, D)

Input: C: the set of all conditional attributes, D: the set of decision attributes.

$Red \leftarrow \{\}; \gamma_{best} = 0; \gamma_{prev} = 0$

repeat

$T \leftarrow Red$

$\gamma_{prev} = \gamma_{best}$

for $\forall a \in (C \setminus Red)$ **do**

if $(\gamma_{Red \cup \{a\}})(D) > \gamma_T D$ **then**

$T \leftarrow Red \cup \{a\}$

end if

end for

$\gamma_{best} = \gamma_T(D)$

$Red \leftarrow T$

until $\gamma_{best} == \gamma_{prev}$

return Red

Illustrating the above defined concept with the help of toy example (as shown in Table 2.1). If the task is to calculate degree of dependency of $A = \{a, b\}$. The similarity relation of feature subset is formed by taking t norm of the constituents features as: $R_A(x) = T_{a \in A} R_a(x)$. Using $T_{a \in A}(x) = min\,\{a_1(x), a_2(x), \ldots a_n(x)\}\, if A = \{a_1(x), a_2(x), \ldots a_n(x)\}$. The equivalence class for decision feature is:
$U \setminus D = \{\{x_1, x_3, x_6\}, \{x_2, x_4, x_5\}\}$.
The degree of dependency of decision feature on singleton sets is calculated using [75] and is given by: $\gamma_{\{a\}}(D) = \frac{1.2}{6}, \gamma_{\{b\}}(D) = \frac{2.4}{6}, \gamma_{\{c\}}(D) = \frac{1.2}{6}, \gamma_{\{d\}}(D) = \frac{1.2}{6}, \gamma_{\{e\}}(D) = \frac{2.2}{6}$ and $\gamma_{\{f\}}(D) = \frac{1.2}{6}$. Since, the value of degree of dependency is highest for the feature subset $\{b\}$, so this feature set is used for finding the potential reduct. Now, iteratively adding features to this feature subset and choosing the one producing highest dependency degree. So, the respective dependencies are: $\gamma_{\{a,b\}}(D) = \frac{2.2}{6}, \gamma_{\{b,c\}}(D) = \frac{2.2}{6}, \gamma_{\{b,d\}}(D) = \frac{2.6}{6}, \gamma_{\{b,e\}}(D) = \frac{2.2}{6}$ and $\gamma_{\{b,f\}}(D) = \frac{2.0}{6}$. The set $\{b, d\}$ has highest dependency, continuing as above. $\gamma_{\{a,b,d\}}(D) = \frac{2.4}{6}, \gamma_{\{b,c,d\}}(D) = \frac{2.2}{6}, \gamma_{\{b,d,e\}}(D) = \frac{2.2}{6}$ and $\gamma_{\{b,d,f\}}(D) = \frac{2.2}{6}$. Since, there is no increase in dependency for these subset than feature subset $\{b, d\}$. The set $\{b, d\}$ is the required reduct.

Table 2.1: Example Dataset

Features / Instances	a	b	c	d	e	f	D
x_1	0.4	0.4	1.0	0.8	0.4	0.2	1
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	0
x_6	0.6	0.6	0.8	0.2	0.8	0.8	1

2.4.1 Related Work

Fuzzy rough set based feature selection has been discussed by many researchers in their work. A discernibiity matrix and similarity relation based fuzzy rough feature selection is proposed in [76]. A different classes ratio based fuzzy rough set model is introduced to diminish the impact of noisy instance in the calculation of lower and upper approximation [90]. The high computational complexity was reduced by neighbourhood approximation and feature grouping in [71]. Further, an accelerator for fuzzy rough feature selection was proposed [119]. A fuzzy rough feature selection for semi supervised data is also introduced where objects are only partially labeled. The generated subsets are only sub optimal with respect to whole dataset [77]. Researchers exploited fuzzy rough set model for identifying the probability of cervical cancer in [86]. A special type of relation known as tolerance relation was employed for feature selection by Jensen et. al. [74]. A method that combines fuzzy c mean clustering and equivalence relation was proposed for feature selection [174]. Also, incremental approach to feature selection is also discussed in [115] using fuzzy rough set. Some of more recent works can be found in [23, 57, 105, 107, 153, 155, 166] that have employed fuzzy rough set model for feature selection.

2.4.2 Limitation of Fuzzy Set

Fuzzy set theory effectively deals with the uncertainty arising in the data, but it has got number of limitations.

- Many of the applications cannot be handled by fuzzy set theory, like the voting problem wherein some of the voters may vote in favour of the item, some may be against the item while some may remain neutral. This could be modeled by having non-membership degree for voters against the item and hesitancy for neutral voters.
- Fuzzy set theory handles vagueness by assigning a membership degree between 0 and 1 but the uncertainty found in judgement and identification cannot be simultaneously handled by single value. Thus, there is a need for some theory that can deal with this uncertainty.

There are number of ways to handle uncertainty arising in information system like vague set and intuitionistic fuzzy set. Vague set assigns membership based on intervals while intuitionistic fuzzy set assigns membership, non-membership and hesitancy. However, vague set more costly than intuitionistic fuzzy set in handling vagueness and Bustice and Burillo [17] showed both of them to be equivalent. Therefore, intuitionistic fuzzy set provides much better way to handle real world ambiguities. So, imprecision could be efficiently modeled by intuitionistic fuzzy set.

2.5 Intuitionistic Fuzzy Set

An intuitionitic fuzzy set [4] is an ordered pair (μ, ν) such that $0 \leq \mu, \nu \leq 1$ and $0 \leq \mu + \nu \leq 1$. For the universe of discourse, an intuitionistic fuzzy set A is given by:

$$A = \{(\mu_A(x), \nu_A(x)) \mid x \in U\} \tag{2.12}$$

where $\mu_A, \nu_A : U \implies [0,1]$ and are called membersip and non-membership degree of an object x respectively.

Some of the properties of intuitionistic fuzzy set are:

- $(\mu_1, \nu_1) = (\mu_2, \nu_2) \leftrightarrow \mu_1 = \mu_2, \nu_1 = \nu_2$
- $(\mu_1, \nu_1) \cap (\mu_2, \nu_2) = (min\{\mu_1, \mu_2\}, max\{\nu_1, \nu_2\})$
- $(\mu_1, \nu_1) \cup (\mu_2, \nu_2) = (max\{\mu_1, \mu_2\}, min\{\nu_1, \nu_2\})$
- The cardinality of intuitionistic fuzzy set is defined as: $\mid A \mid = \sum_{x \in U} \frac{1+\mu_A(x)-\nu_A(x)}{2}$
- An intuitionistic fuzzy relation R is an equivalence relation iff:
    1. Reflexive: $\mu_R(x,x) = 1$ and $\nu_R(x,x) = 0$
    2. Symmetric: $\mu_R(x,y) = \mu_R(y,x)$ and $\nu_R(x,y) = \nu_R(y,x), \forall x, y \in U$
    3. Transitive: $\mu_R(x,z) \geq \vee(\mu_R(x,y) \wedge \mu_R(y,z))$ and $\nu_R(x,z) \leq \wedge(\nu_R(x,y) \vee \nu_R(y,z)), \forall x, y, z \in U$. An intuitionistic fuzzy relation is tolerance relation if its reflexive and symmetric.
- Some of the common intuitionistic fuzzy t norms are: $T_M(x,y) = (min(x_1, y_1), max(x_2, y_2))$ and $T_W = (max(0, x_1+y_1-1), min(1, x_2+y_2))$, where $x = (x_1, x_2), y = (y_1, y_2)$.
- $I_M(x,y) = (max(x_2, y_1), min(x_1, y_2))$ and $I_W(x,y) = (min(1, x_2+y_1), max(0, x_2+y_2-1))$ are some of the common intuitionistic fuzzy implicators.

2.5.1 Intuitionistic Fuzzy Rough Set Theory

Integrating rough and intuitionistic fuzzy set theory provides a better or more precise way of handling uncertainty and vagueness occurring in real valued datasets.

Employing the above defined terms, intuitionistic fuzzy lower and upper approximations are defined as:

$$R \downarrow_A X(x) = inf_{y \in U} I(R(x,y), X(y)) \tag{2.13}$$

$$R \uparrow_A X(x) = sup_{y \in U} T(R(x,y), X(y)) \tag{2.14}$$

where I and T are intuitionistic fuzzy implicator and t norm respectively. The pair $(R \downarrow_A X, R \uparrow_A X)$ is called intuitionistic fuzzy rough set. These approximations are employed in calculating degree of dependency of feature subset and thereby applied for feature selection.

2.5.2 Related Work

Intuitionistic fuzzy rough set theory is an emerging model that has been applied by few researchers for feature selection. A genetic algorithm for feature selection was employed by Lu et. al. [99]. A combination of information entropy with intuitionistic fuzzy rough set was used for attribute reduction [25]. The structure and properties of intuitionistic fuzzy rough set was discussed and employed for feature selection in [44]. An intuitionistic fuzzy rough set model based on distance function and discernibility matrix was applied for feature selection by Huang et. al. [64] and Zhang [173] respectively. However, none of them discussed intuitionistic fuzzy rough set model based on dependency function except few works like [142, 143], etc.

In the current thesis, various methods for feature selection employing fuzzy and intuitionistic fuzzy rough set models are discussed while addressing the issues like noise, incompleteness, outliers, etc arising in the dataset.

2.6 Summary

This chapter dive into the mathematical details of theories that are used to solve dimensionality problem i.e. feature selection. The notion of rough set theory along wih its application in feature selection is given. However, the need for data discretization to apply rough set based feature selection may lead to information loss. Fuzzy set theory, an extension of rough set theory, associates a membership value with each object. The hybridization of fuzzy and rough set effectively handles vagueness and indiscernibility arising in the data along with overcoming the need for data discritization. Fuzzy rough set theory has therefore got lot of applications in information retrieval, feature selection, etc. Fuzzy rough set based feature selection along with related work in this field is given in this chapter.

A step advancement in fuzzy set theory, intuitionistic fuzzy set theory employs membership and non-membership value to show the belongingness of an object to a set. Intuitionistic fuzzy rough set theory is therefore being effectively applied for feature selection.

****