Rough Sets in Data Mining and Spatial Computing

Sonajhaira Minz
Professor
School of Computer & Systems Sciences,
Jawaharlal Nehru University,
New Delhi 110 067 INDIA
http://www.jnu.ac.in/FacultyStaff/Showrofile.asp?SendUserName=sminz

Outline

PART –I

- Rough Set Theory Applied for Data Mining
- Data Mining using Rough Sets

PART - II

- Remote sensing data classification and Rough Sets
- Spatial Computing

Background

Data Mining & Rough Set Theory – Classification, Clustering:

- Rajni Jain, Girish Kumar Singh,
- Arun, Ashish, Puran, Yashpal, Sunil, Kalicharan, Surender, Mary, Ibrahim, Rangbahadur,

Multi Agent Systems for Scheduling Problems: Ahmad Balid

Privacy Preserving Data Mining: Fuad Al Yamiri

Time-Series Analysis: Ibrahim Abu Ghali

Multi-view Ensemble Learning: Vipin Kumar

PART-I

Applications of Rough Set Concepts for Data Analysis

- Real Data
- Data Models
- Issues in dimension reduction

Real Data

- Vector/multidimensional representation
  - Numeric values of the attributes
- Mostly dirty
- Imprecise
- Not essentially large

Data Models

- Patterns that describe the data
- Concepts, Class Descriptions, Clusters, Association Rules

Data Mining: Machine Learning

Decision Tree Induction: Attribute Selection for test at an interior node

Problems:

- Conceptually/quantitatively - Correlated Attributes
- Over fitted training samples
- Attribute selection error

Rough Sets: Real Data Experience

- Reduct
  - Types
  - Data reduction
  - Concept description: Cluster analysis
- POS
  - Discretization
- Granulation Tree
  - Association Rule mining
  - Clustering

Types of Reduct

- (Decision) Relative Reduct
- Core = $\cap$ R_i, Reduct = $\{R_i\}$ i= 1,2,…
- Variable Precision RS: Approximate Core
- Dynamic Reduct

Decision relative Reduct

Reduct in comparison to simple diversity index, entropy, information gain, gain ratio, GINI index

Complexity -

- reduct computation: $O(m^2 n \log n)$
- tree induction: $O(m\ n \log n)$
- DT induction (with tree prunning): $O(m\ n \log n) + O(n (\log n)^2)$
- RDT: $[O(m^2 n \log n) + O(|R|\ n \log n) + O(n (\log n)^2)$

Variable Precision RS: Approximate Core

- $\mathbb{R} = \{R_1, R_2, \ldots, R_n\}$ Set of all reducts
  Core = $R_1 \cap R_2 \cap \ldots \cap R_n$
- $\mathbb{A} = R_1 \cup R_2 \cup \ldots \cup R_n$
  $|\mathbb{A}| = m$
- core $\subseteq$ core$_\alpha \subseteq \mathbb{A}$

Approximate core - (core$_\alpha$: α') where,

α‘ degree of approximation of the set core$_\alpha$ a superset of core.

Approximate Core Cont.

Approximate core a superset of core.

C_1: ($core_\alpha$: 1) ≡ core, say $\{a_1\}$

C_2: ($core_\alpha$: 0.98) ≡ $\{a_1, a_5\}$

C_3: ($core_\alpha$: 0.95) ≡ $\{a_1, a_2, a_5, a_7\}$

C_4: (: ($core_\alpha$: 0) ≡ $\{a_i: a_i \notin \mathbb{A}\}$

- The additional attribute added to core from $\mathbb{A}$ is selected from smallest reduct in $\mathbb{R}$.
- Approximate core instead of reduct RDT $core_\alpha$: α

Dynamic Reduct

Dynamic RDT Dynamic reduct based decision tree for handling noise.

Information System T = (U, A, V, F), where A = C $\cup$ D

C: Conditional attributes

D: Decision attribute

Let power set of T be $\mathcal{P}$(T). A subtable π of T is a member of $\mathcal{P}$(T) with respect to the attribute set A.

Dynamic Reduct

Let a subtable π corresponding the attribute set $\subseteq \mathcal{P}(\mathrm{T})$.

DR(T, π): Dynamic Reduct of information T.

RED(T, D): Decision relative reduct of T.

RED_{π}(B, D): Decision relative reduct of B with respect to subatable π.

DR(T, π) = $\cap_{\pi \in \mathcal{P}(\mathrm{T})} \mathrm{RED}_{\pi}$(B, D)

Discretization

- Boolean reasoning
- Using positive region POS of class

A supervised (labelled data) grouping- maximizing the cardinality of $\mathrm{POS}_{a_i}(C_p)$ for numeric values of attribute a_i, with respect to decision class C_p using a rough membership function f_{a_i,c_p}

Discretization

Rough membership Function $f_{a_i,c_p}: V_{a_i} \rightarrow R$

$$\arg max_{x \in I \subseteq V_{a_i}} f_{a_i,c_p}(x) = \left\{ x \left| \frac{card(\underline{a_{i,I}} X_{C_p})}{card(X_{a_i,I})} \right. \right\}$$

Where,

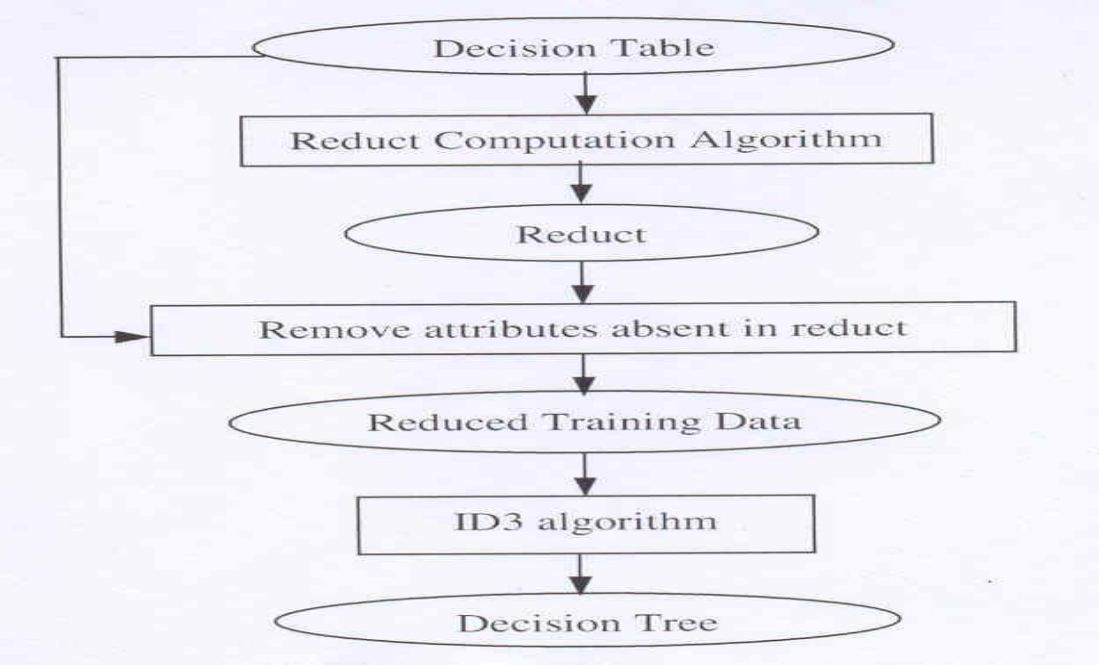
$X_{a_i,I} = \{x |\ x \in U, a_i(x) \in I, I \subseteq V_{a_i}\}$ and,

$\underline{a_{i,I}} X_{C_p} = \{ x |\ a_i(x) \in I, I \subseteq V_{a_i}, D(x) = c_P\}$

Also, $POS_{a_i}(D) = \cup \{\underline{a_i} X : X \in [x]_D\}$

Reduct: RDT for Classification

Achieve Dimension Reduction using reducts



RDT Experiment

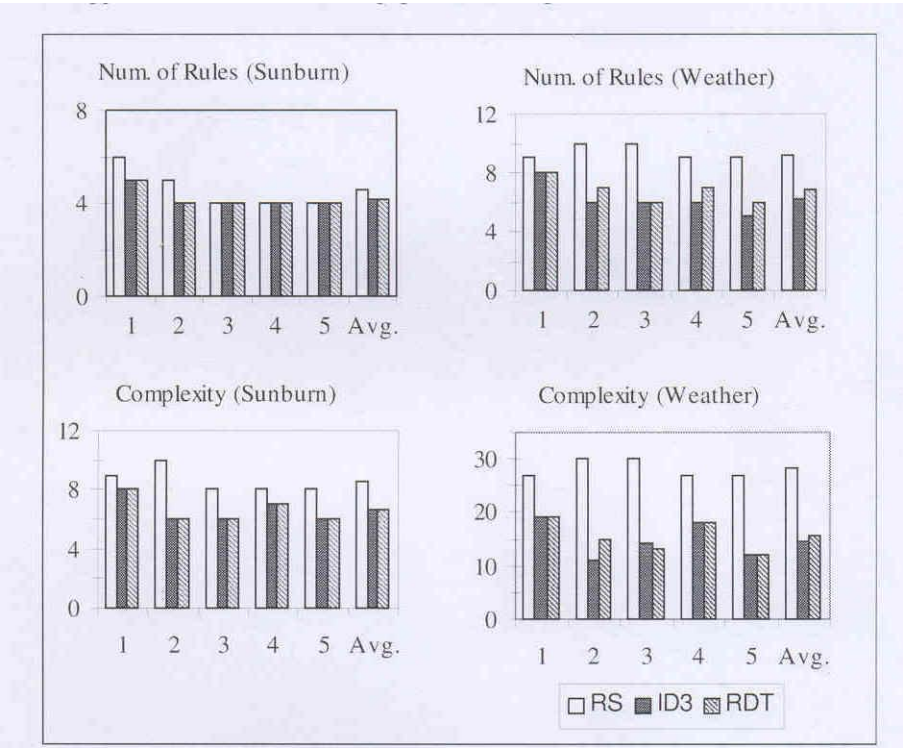


Fig. 4. Comparison of RS, ID3, RDT algorithms w.r.t. number of rules, complexity for training:test::70:30 of Sunburn and Weather

RDT Experiment

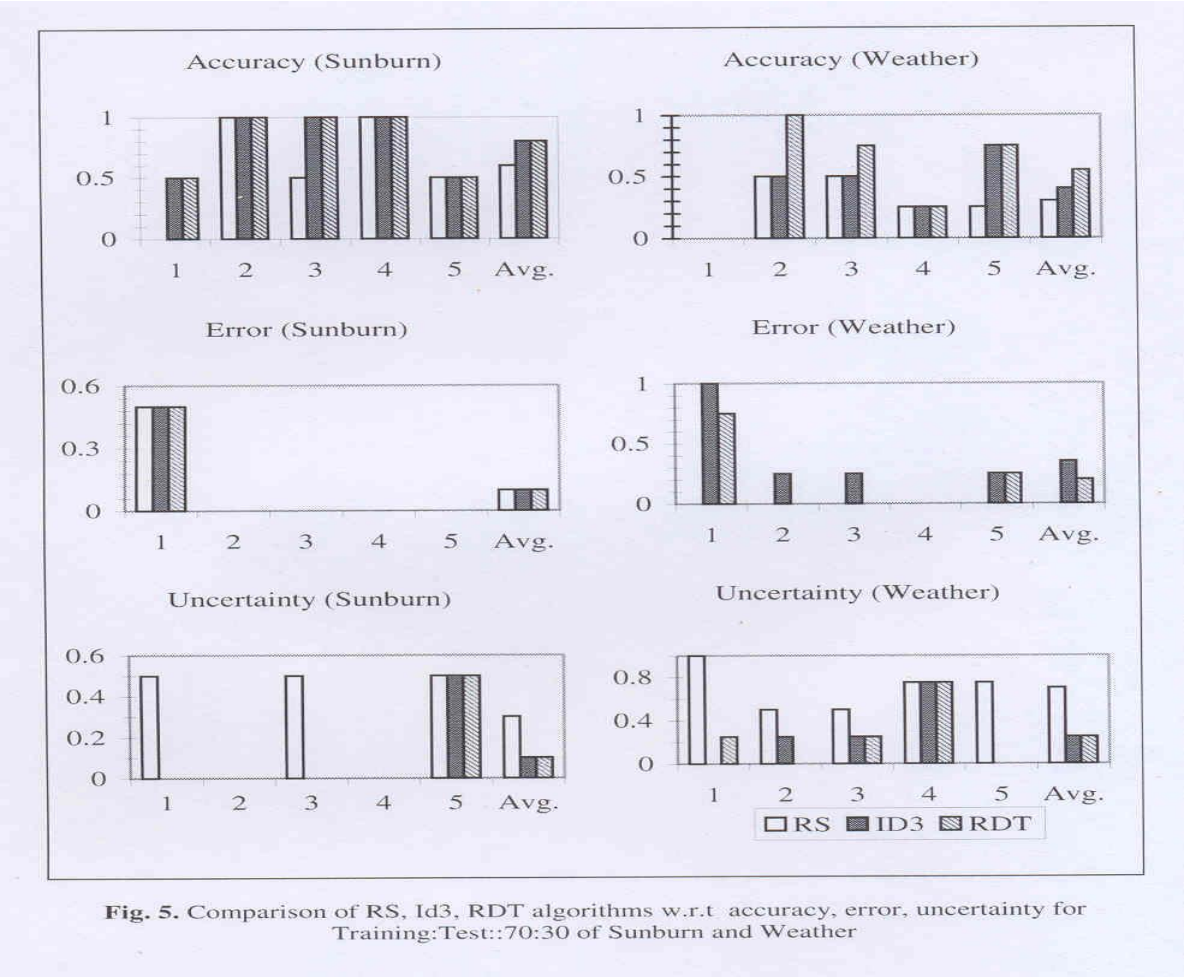


Fig. 5. Comparison of RS, Id3, RDT algorithms w.r.t accuracy, error, uncertainty for Training:Test::70:30 of Sunburn and Weather

Reduct in other models

- Applied to Covering algorithm
- Applied to Cluster analysis

Reduct based Covering Algorithm

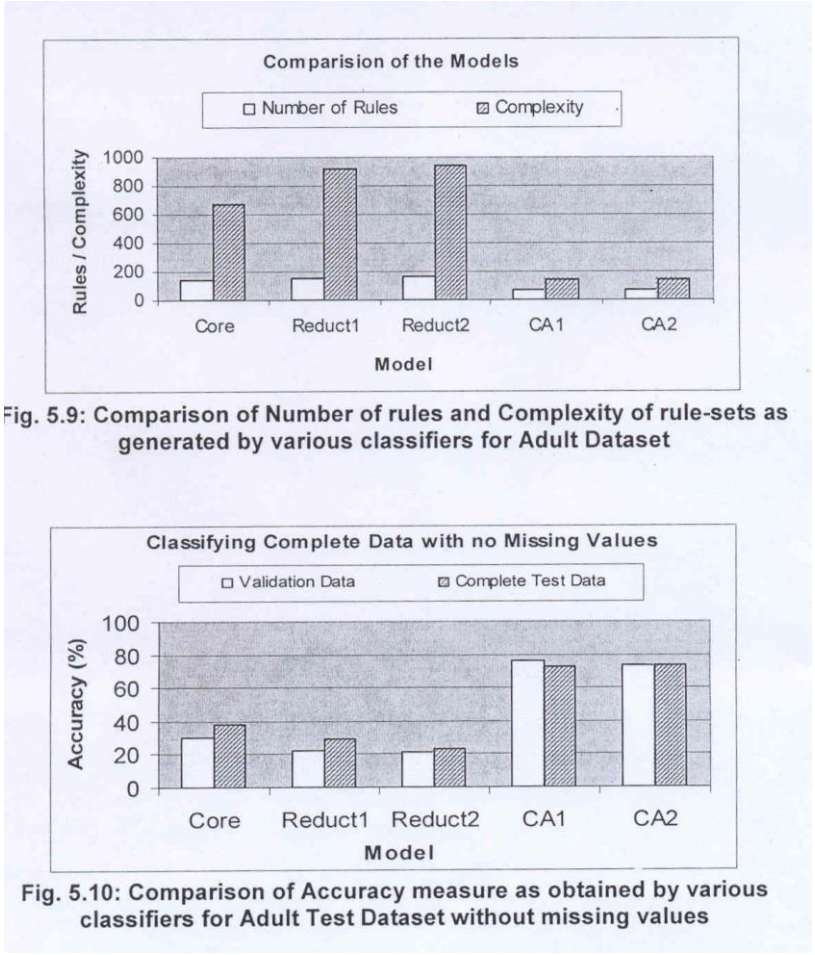


Fig. 5.9: Comparison of Number of rules and Complexity of rule-sets as generated by various classifiers for Adult Dataset

Fig. 5.10: Comparison of Accuracy measure as obtained by various classifiers for Adult Test Dataset without missing values

Reduct based Covering Algorithm

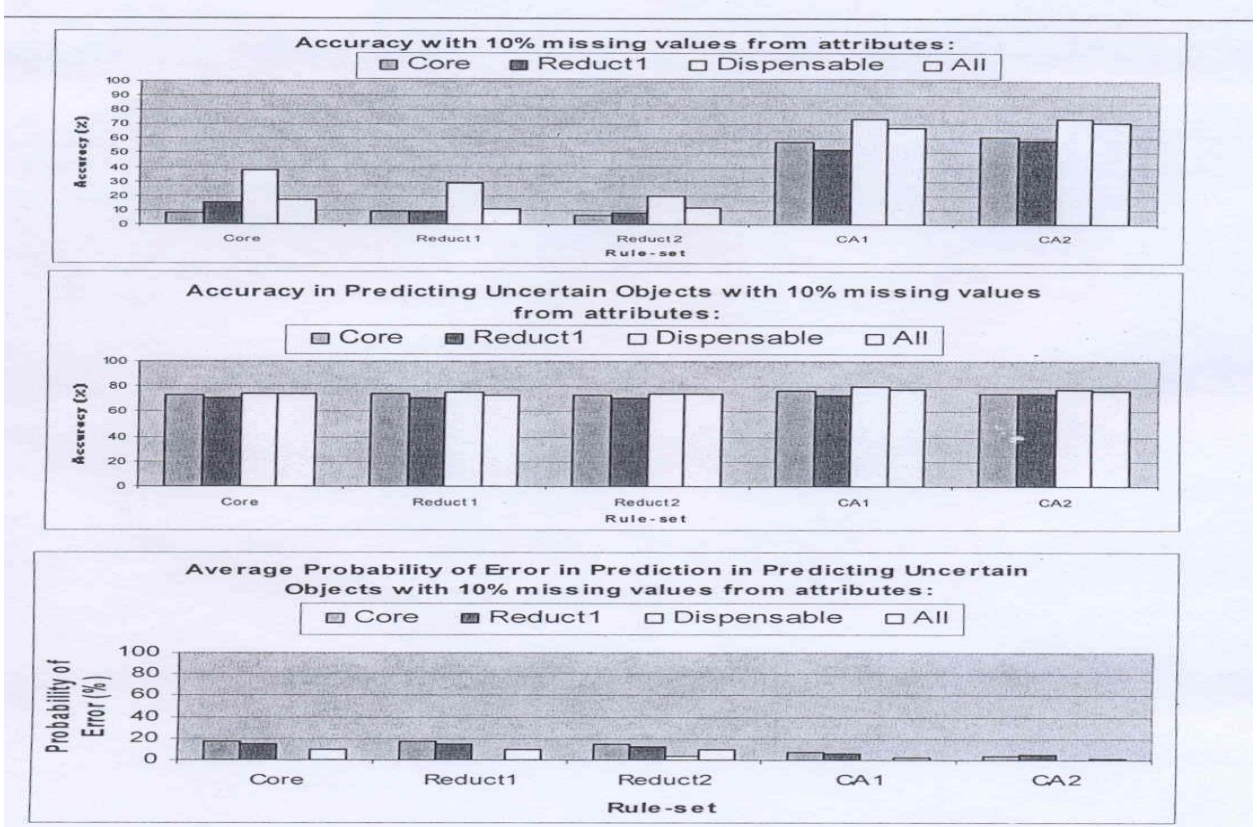


Figure 5.13 Comparison of accuracy and probability of error in predicting uncertain objects as obtained by using various classifiers on Adult Test Dataset with 10% missing values from four different categories of attributes

Reduct based Covering Algorithm

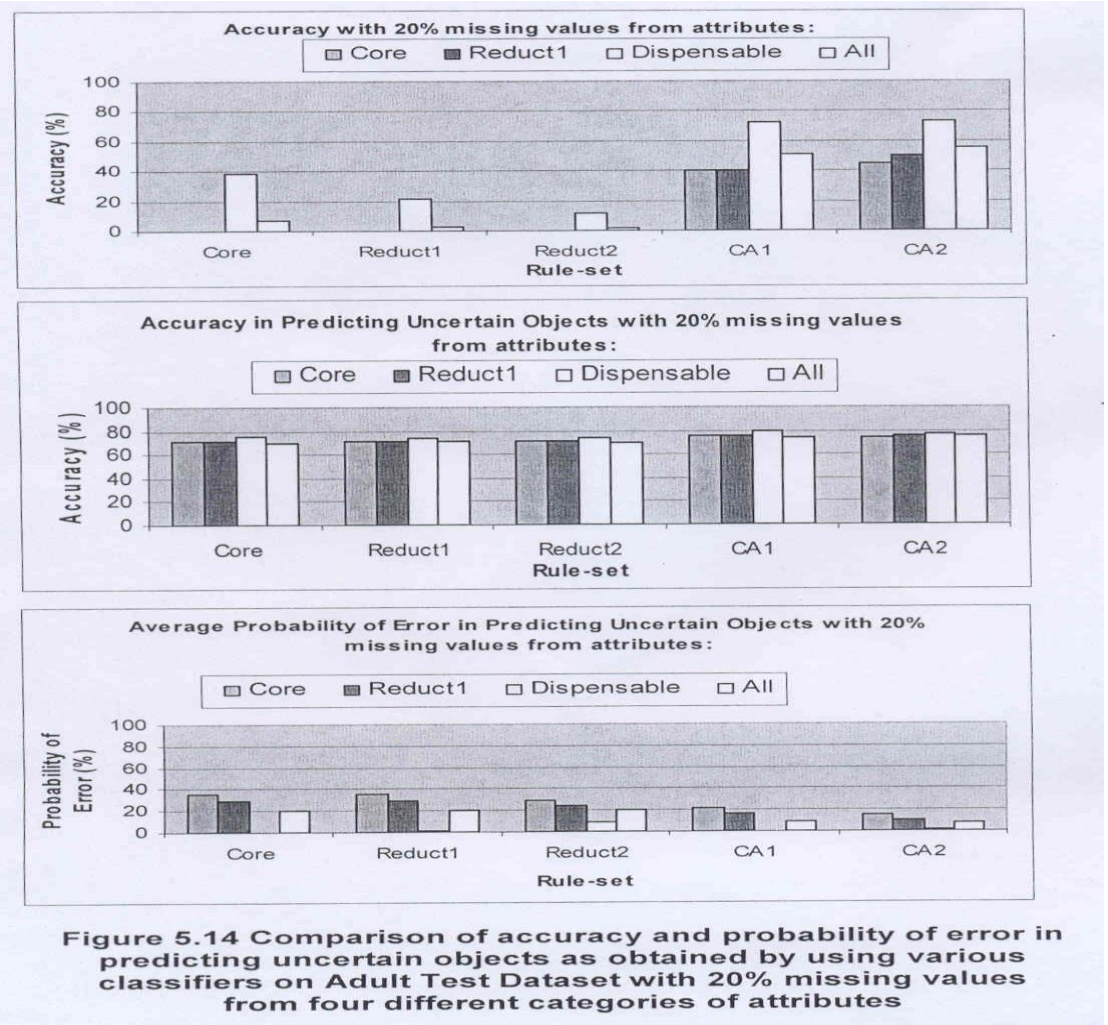


Figure 5.14 Comparison of accuracy and probability of error in predicting uncertain objects as obtained by using various classifiers on Adult Test Dataset with 20% missing values from four different categories of attributes

Reduct based Cluster Analysis

- Iris Data

Table 6.3: Existence of clusters obtained by DBSCAN and AHC

Clustering Algorithm	No. of Cluster	Reduct if any	Significance of attributes
DBSCAN	4	{SL,SW,PL}, {SL,SW,PW}, {SW,PL,PW }	SL= 0.13, SW=0.17, PL=0.86, PW=0.69
AHC	4	{PL, PW}	SL= 0.19, SW=0.21, PL=0.8, PW=0.78

Reduct based Cluster Analysis

Table 6.5: Difference of clusters obtained by AHC

Pair wise Clusters for analysis		Reduct if any	Significance of attributes
First cluster	Second Cluster		
1	2	{PL}, {PW}	SL=0.73, SW=0.44, PL=1, PW=1
1	3	{PL}, {PW}	SL=0.39, SW=0.93, PL=1, PW=1
1	4	{PL}, {PW}	SL=0.9, SW=0.32, PL=1, PW=1
2	3	{PL}	SL=0.63, SW=0.6, PL=1, PW=0.39
2	4	{PW}	SL=0.24, SW=0.09, PL=0.66, PW=1
3	4	{PL}, {PW}	SL=0.84, SW=0.72, PL=1, PW=1

Variable Precision RS: Approximate Core

- $\mathbb{R} = \{R_1, R_2, \ldots, R_n\}$ Set of all reducts
  Core = $R_1 \cap R_2 \cap \ldots \cap R_n$
- $\mathbb{A} = R_1 \cup R_2 \cup \ldots \cup R_n$
  $|\mathbb{A}| = m$
- core $\subseteq$ core$_\alpha \subseteq \mathbb{A}$

Approximate core - (core$_\alpha$: α') where,

α‘ is the degree of approximation of the set core$_\alpha$ a superset of core.

Approximate Core Cont.

Approximate core a superset of core.

C_1: $(core_\alpha: 1) \equiv$ core, say $\{a_1\}$

C_2: $(core_\alpha: 0.98) \equiv \{a_1, a_5\}$

C_3: $(core_\alpha: 0.95) \equiv \{a_1, a_2, a_5, a_7\}$

C_4: $(core_\alpha: 0) \equiv \{a_i: a_i \notin \mathbb{A}\}$

- The additional attribute added to core from $\mathbb{A}$ is selected from smallest reduct in $\mathbb{R}$.
- Approximate core instead of reduct RDT $core_\alpha$: α

Approximate Core: RDT $core_{\alpha}$: α

Table 1: Learning schemes and their descriptions used for Forest cover type dataset

ALGORITHM	DESCRIPTION
RS	Classical Rough set approach with full discernibility decision relative reduct
CJU	Continuous data, J4.8 algorithm, Unpruned - Java implementation of C4.5
CJP	Continuous data, J4.8 Algorithm, Pruned - Java implementation of C4.5
RJU	Discretized, filtered using Reducts, J4.8 unpruned
RJP	Discretized, filtered using Reducts, J4.8 pruned
RDTGA-smallest	Discretized, filtered using smallest reduct from the population, ID3
RJUGA-smallest	Discretized, filtered using smallest reduct from the population, J4.8 unpruned
RJPGA-smallest	Discretized, filtered using smallest reduct from the population, J4.8, pruned
$RJUcore_{\alpha:c}$	Discretized, filtered using approx. core with $\alpha \geq c$, J4.8, unpruned
$RJPcore_{\alpha:c}$	Discretized, filtered using approx. core with $\alpha \geq c$, J4.8, pruned

RDT core$_\alpha$: α

Table 2: Frequencies of attributes in population of reducts from 5 randomly selected samples of training data for approximate core identification

#1	w1	#2	w2	#3	w3	#4	w4	#5	w5
1	1	1	1	1	1	1	1	1	1
2	1	2	1	2	1	2	1	2	1
3	1	3	1	3	1	3	1	3	1
4	1	4	1	4	1	4	1	4	1
5	1	5	1	5	1	5	1	5	1
6	1	6	1	6	1	6	1	6	1
7	1	7	1	7	1	7	1	7	1
8	1	8	1	8	1	8	1	8	1
9	1	9	1	9	1	9	1	9	1
10	1	10	1	10	1	10	1	10	1
37	0.99	52	0.99	26	1	24	1	46	1
45	0.96	25	0.97	43	0.97	43	0.99	47	0.99
44	0.96	26	0.97	20	0.97	36	0.98	25	0.98
52	0.96	43	0.97	30	0.96	52	0.98	37	0.98
34	0.95	45	0.97	36	0.96	47	0.97	34	0.97
36	0.96	46	0.97	37	0.96	26	0.97	24	0.96
46	0.96	36	0.96	53	0.95	34	0.97	43	0.96
47	0.94	47	0.96	24	0.93	25	0.95	26	0.93
30	0.93	37	0.95	45	0.93	37	0.95	13	0.90
43	0.93	53	0.95	16	0.90	38	0.91	36	0.89
26	0.91	28	0.92	34	0.90	45	0.91	33	0.87
16	0.89	44	0.92	46	0.90	46	0.91	38	0.84
24	0.89	24	0.91	38	0.89	16	0.88	18	0.82
20	0.87	16	0.90	27	0.87	33	0.85	20	0.82

RDT core$_\alpha$: α

Table 5: Comparison of CS for the Dataset

Algorithm	A (%)	Att	R (10^3)	S (10^4)	CS
RS	13.4	29	24	70	0.04
CJU	82.3	53	28	73	0.21
CJP	82.5	51	2	55	0.21
RDTGA	74.1	29	11	5	0.19
RJUGA	79.8	29	30	29	0.21
RJPGA	78.4	29	78	6	0.20
RJUcore$_{\alpha:1}$	77.6	10	35	18	0.22
RJPcore$_{\alpha:1}$	75.6	10	6	2	0.21

Notes:
A: accuracy of the model
Att: number of attributes used in the classifier
R: number of rules in the classifier
S: number of selectors in the classifier

Table 6: Comparison of accuracy with previously reported results for Forest cover type dataset

MODEL	A
Back Propagation	70%
Linear Discriminant Analysis	58%
SVM	71%
SVM modified for unrepresentative class	73.41%
C5	83.7%
CHAID	72.7%
CART	68.9%
XCS	66.9%
CJU	82.3%
CJP	82.6%
RDTGA-smallest	74.1%
RJUGA-smallest	79.8%
RJPGA-smallest	78.5%
RJUcore$_\alpha$:1	77.0%
RJPcore$_\alpha$:1	75.0%

Dynamic Reduct

Dynamic RDT Dynamic reduct based decision tree for handling noise.

Information System T = (U, A, V, F), where A = C $\cup$ D

C: Conditional attributes

D: Decision attribute

Let power set of T be $\mathcal{P}$(T). A subtable π of T is a member of $\mathcal{P}$(T) with respect to the attribute set A.

Dynamic Reduct

Let a subtable π corresponding the attribute set $\subseteq \mathcal{P}(\mathrm{T})$.

DR(T, π): Dynamic Reduct of information T.

RED(T, D): Decision relative reduct of T.

RED_{π}(B, D): Decision relative reduct of B with respect to subatable π.

DR(T, π) = $\cap_{\pi \in \mathcal{P}(\mathrm{T})} \mathrm{RED}_{\pi}$(B, D)

Dynamic Reduct based RDT: Nutrition Dataset (Real)

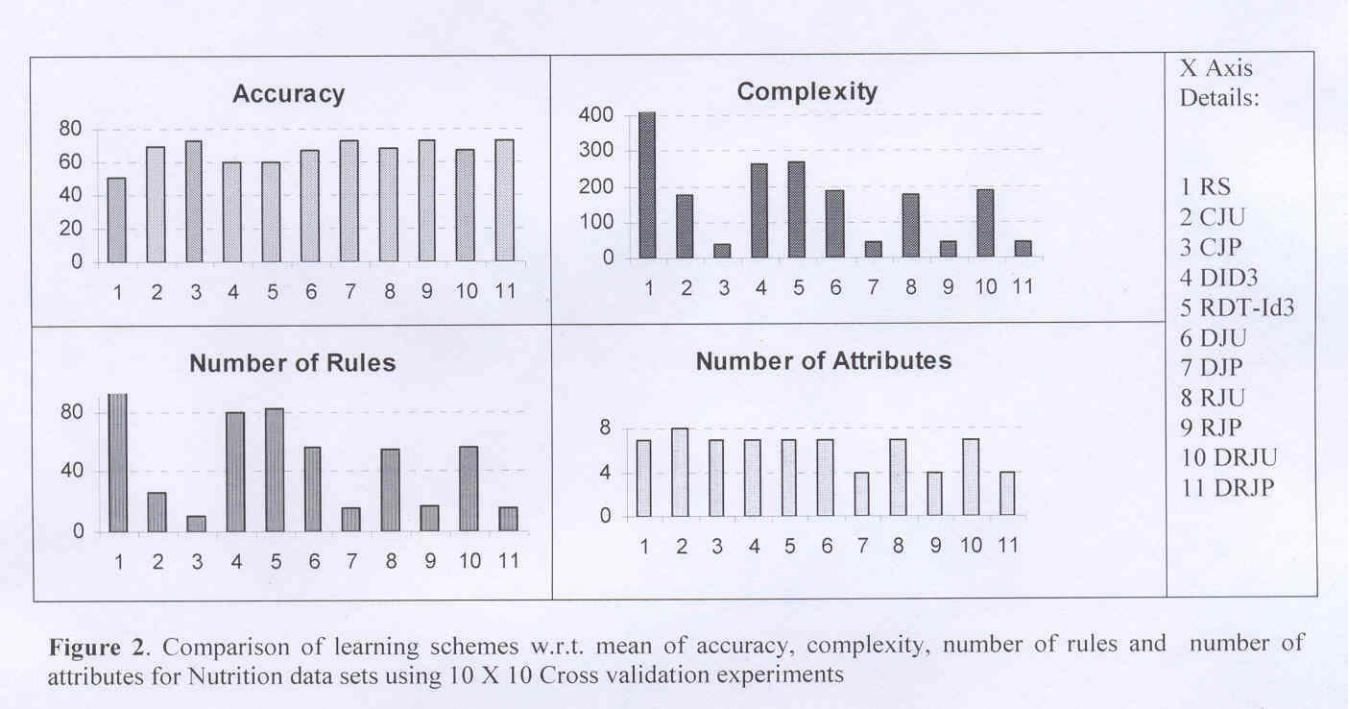


Figure 2. Comparison of learning schemes w.r.t. mean of accuracy, complexity, number of rules and number of attributes for Nutrition data sets using 10 X 10 Cross validation experiments

Discretization

- Boolean reasoning
- Using positive region POS of class

A supervised (labelled data) grouping- maximizing the cardinality of $\mathrm{POS}_{a_i}(C_p)$ for numeric values of attribute a_i, with respect to decision class C_p using a rough membership function f_{a_i,c_p}

Discretization

Rough membership Function $f_{a_i,c_p}: V_{a_i} \rightarrow R$

$$\arg max_{x \in I \subseteq V_{a_i}} f_{a_i,c_p}(x) = \left\{ x \left| \frac{card(\underline{a_{i,I}} X_{C_p})}{card(X_{a_i,I})} \right. \right\}$$

Where,

$X_{a_i,I} = \{x |\ x \in U, a_i(x) \in I, I \subseteq V_{a_i}\}$ and,

$\underline{a_{i,I}} X_{C_p} = \{ x |\ a_i(x) \in I, I \subseteq V_{a_i}, D(x) = c_P\}$

Also, $POS_{a_i}(D) = \cup \{\underline{a_i} X : X \in [x]_D\}$

Discretization using POS

Table 3.2: Data Set Description

Properties	Datasets			
	Iris	Ion	Hea	Pid
No. of Examples	150	351	270	768
No. of Classes	3	2	2	2
No. of Attributes	4	34	13	8
No. of Cont. Attributes	4	32	6	8
All-Cont/Mix-mode	All-Cont	Mix-mode	Mix-mode	All-Cont

Discretization using POS

- Result-1

Table 3.3: Comparison of the Eight Discretization Schemes for Labeled Data based on CAIR Value

Discretization Method	Datasets			
	Iris	**Ion**	**Hea**	**Pid**
Equal Width	0.40	0.098	0.087	0.058
Equal frequency	0.41	0.095	0.079	0.052
Patterson-Niblett	0.35	0.192	0.088	0.052
IEM	0.52	0.193	0.118	0.079
Max. Entropy	0.30	0.100	0.081	0.048
CADD	0.51	0.130	0.098	0.057
CAIM	0.54	0.168	**0.138**	0.084
Proposed Method	**0.56**	**0.237**	**0.128**	**0.107**

Discretization using POS

- Result-2

Table 3.4: Comparison of the Eight Discretization Schemes for Labeled Data based on Number of Intervals

Discretization Method	Datasets			
	Iris	**Ion**	**Hea**	**Pid**
Equal Width	16	640	56	106
Equal frequency	16	640	56	106
Patterson-Niblett	48	384	48	62
IEM	**12**	113	**10**	**17**
Max. Entropy	16	572	56	97
CADD	16	536	55	96
CAIM	**12**	**64**	12	**16**
Proposed Method	**12**	**85**	**11**	**33**

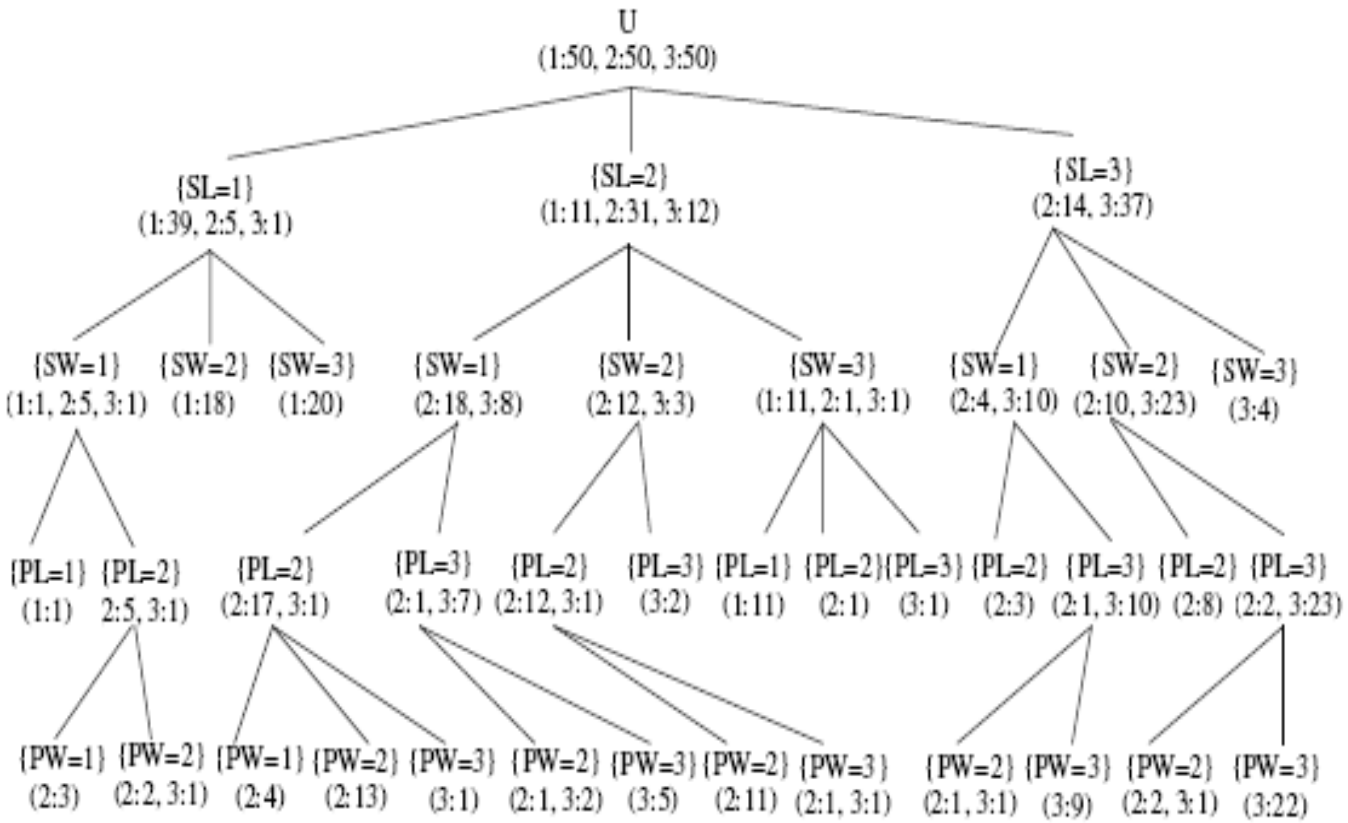
Discretization using POS

- Result-3

Table 3.5: Comparison of the Discretization by proposed scheme for labeled and unlabeled data

Evaluation Parameter	Discretization Scheme	Datasets	
		Iris	Pid
CAIR Value	Labeled	**0.56**	**0.107**
	Unlabeled	**0.53**	**0.105**
Numbers of Intervals	Labeled	**12**	**33**
	Unlabeled	**12**	**37**

Granulation Tree



Granulation Tree

Table 4.1: Granulation Result

Result	Datasets			
	Iris	Ion	Hea	Pid
1. No. of Granules	23	89	153	209
2. No. of Fine Granules	18	88	151	145
3. No. of Coarse Granules	5	1	2	64
4. No. of object causing coarseness	5	1	2	128

Clustering using Granules

Data Description

Table 5.1 Data Set Description

Properties	Datasets				
	Iris	Solar Flare	Soybean	Nursery	Weather
Number of Instances	150	1066	307	3240	14
Number of Attributes	5	13	35	8	5
Missing value	None	None	Yes	None	None
Type of attribute	Numeric	Nominal	Nominal	Nominal	Nominal

Clustering using Granules

Class	1	2	3
sepal length			
1	(47/62):0.758065	(11/98):0.112245	(1/108):0.009259
2	(3/118):0.025424	(36/85):0.423529	(32/89):0.359551
3	(0/70):0.000000	(3/67):0.044776	(17/53):0.320755
sepal width			
1	(36/102):0.352941	(23/115):0.200000	(29/109):0.266055
2	(13/52):0.250000	(0/65):0.000000	(2/63):0.031746
3	(1/96):0.010417	(27/70):0.385714	(19/78):0.243590
petal length			
1	(50/50):1.000000	(0/100):0.000000	(0/100):0.000000
2	(0/104):0.000000	(48/56):0.857143	(6/98):0.061224
3	(0/96):0.000000	(2/94):0.021277	(44/52):0.846154
petal width			
1	(49/50):0.980000	(0/99):0.000000	(0/99):0.000000
2	(1/57):0.017544	(7/51):0.137255	(0/58):0.000000
3	(0/91):0.000000	(38/53):0.716981	(3/88):0.034091
4	(0/102):0.000000	(5/97):0.051546	(47/55):0.854545

AVERAGE GRANULES QUALITY: 0.20

Figure 5.1Attribute significance matrix for attribute selection

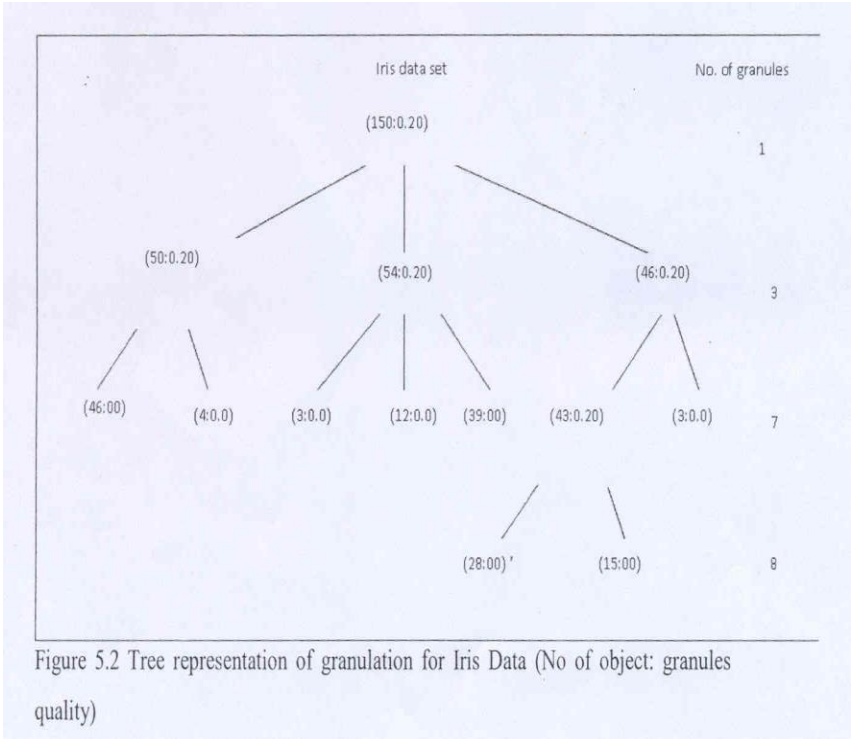
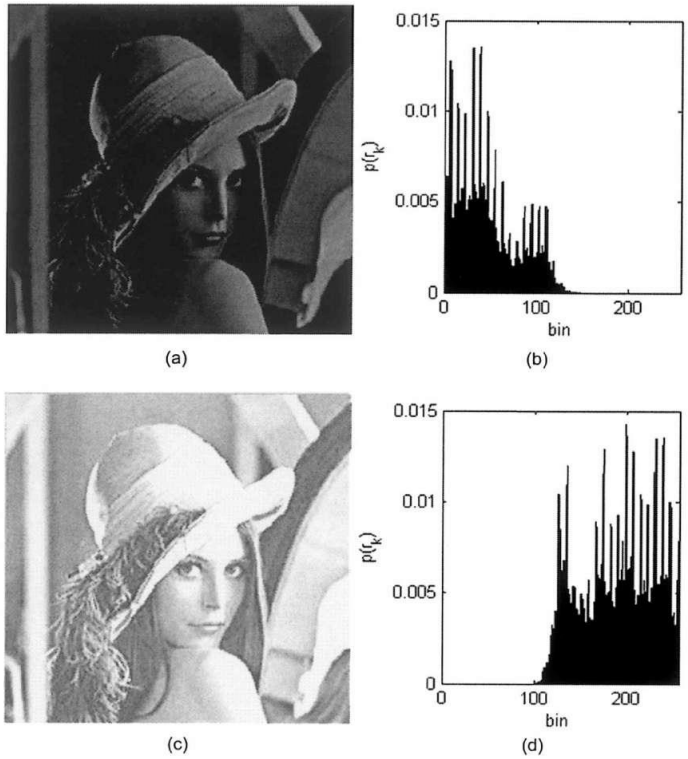


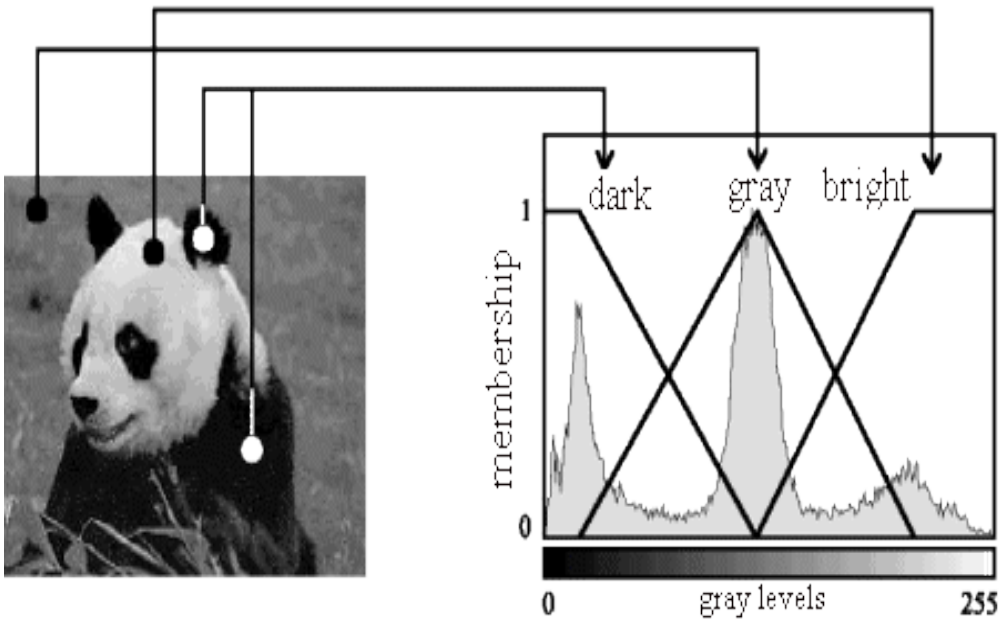
Figure 5.2 Tree representation of granulation for Iris Data (No of object: granules quality)

Histon for Image processing using Rough Sets

- Let an image I of size M×N of L intensity levels $g_{i,j} \in [0, L-1]$.
- The histogram indicates the number of pixels along Y-axis corresponding intensity values 0-255 along X-axis, indicate lower approximation or positive region.
- For segmentation an intensity to identify different concepts in a data

Image Processing: Histogram





Rough Set: Histon

Color images have three histograms each for R, G and B intensity values. For i $\in$ {R, G, B} and 0≤g< L a histogram is

h(g) = $\sum_{m=1}^{M}\sum_{n=1}^{N}\delta(I(m,n)-g)$; Where δ(.) is the Dirac impulse function.

Histon H(g)

H(g) = $\sum_{m=1}^{M}\sum_{n=1}^{N}(1+X(m,n))\delta(I(m,n)-g)$

Where, $X(m,n)=\begin{cases}1\ d_\tau(m,n)<threshhold\\0\ otherwise\end{cases}$, $d_\tau(m,n)$ indicates difference in intensities of a pixel (m,n) with its neighbours

Part -II

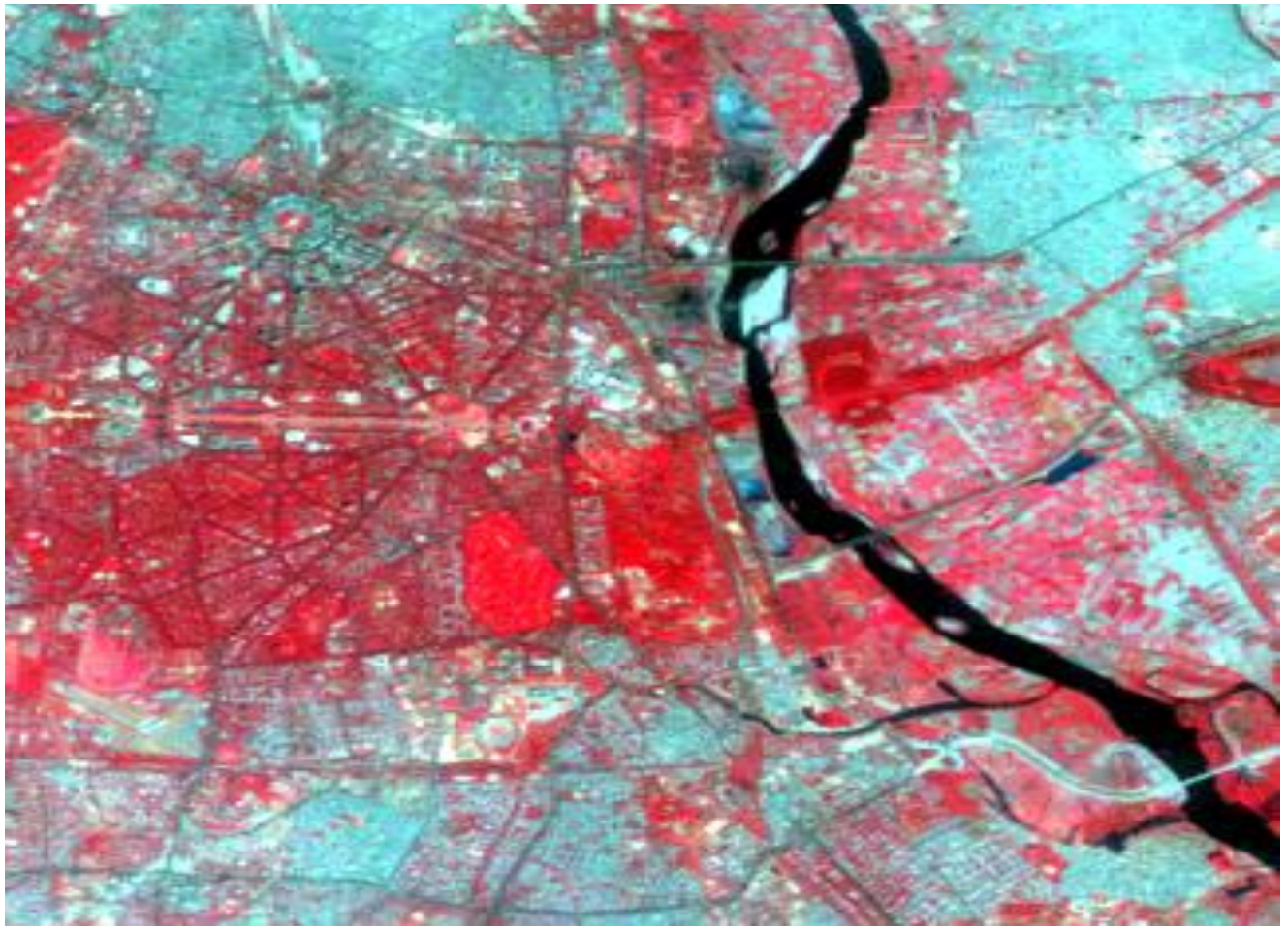
- Spatial Data
- Spatial Computing

Spatial & Remote Sensing Data: The Team

Mining Spatial & Remote Sensing Data (Big Data Analytics)

- Spatial Data Mining: A Machine Learning Approach, Dr. Anshu Dixit Ph.D. 2012 (IASRI)
- "*Change Detection Using Unsupervised Learning Algorithms for Delhi, India*," Asian Journal of Geoinformatics, Vol 13, No. 4, 12-15, 2013, Hemant Kr. Aggarwal, M.Tech. 2013 (Ph.D. IIITD)
- Active Learning for Semi-supervised Classification in Hyperspectral Remote Sensing Images, Monoj Pradhan (Ph.D)
- Semi-Supervised Classification, Prem Shankar (Ph.D)
- Content-based Classification, Saroj Kumar Sahu (Ph.D)
- **Rough Set Based Extreme Learning Machine for Hyper Spectral Data Classification, Ankit Malviya (M.Tech)**
- Comparing Decision Tree and Markov Random Field based Classification for Spatial Data, Mahedra Gupta (M.Tech)

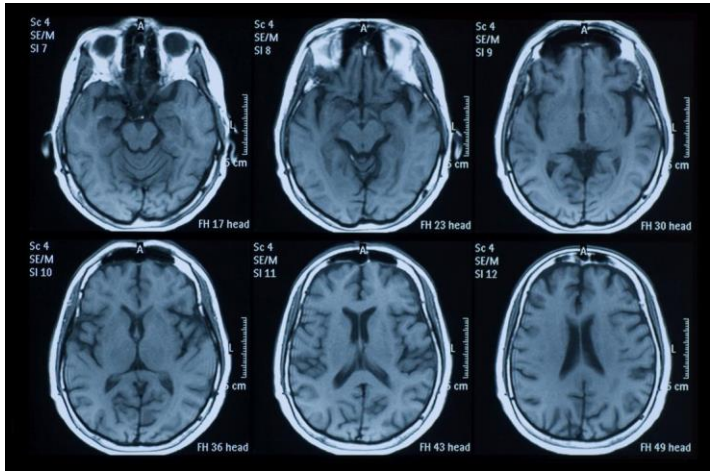
Spatial Data: Satellite Image

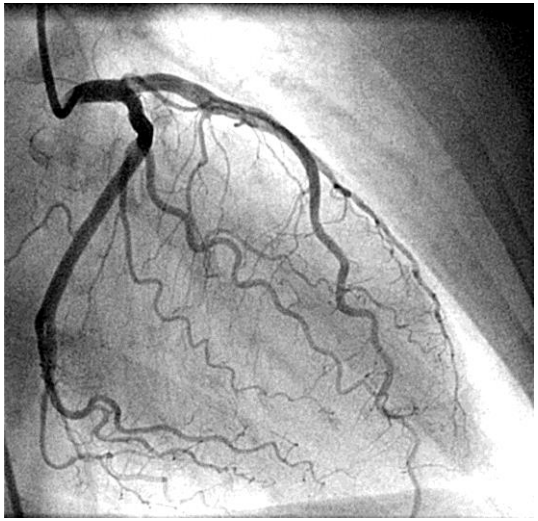


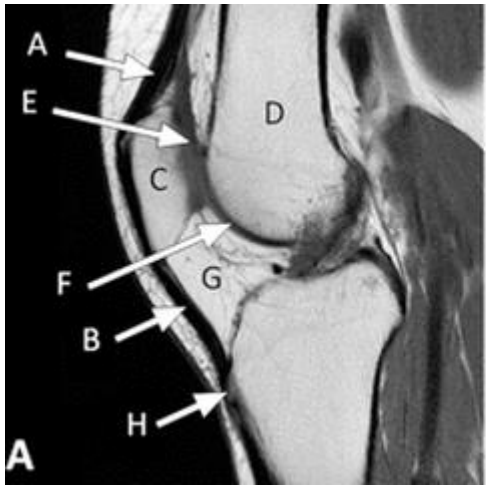
Spatial Data: Multidimensinal

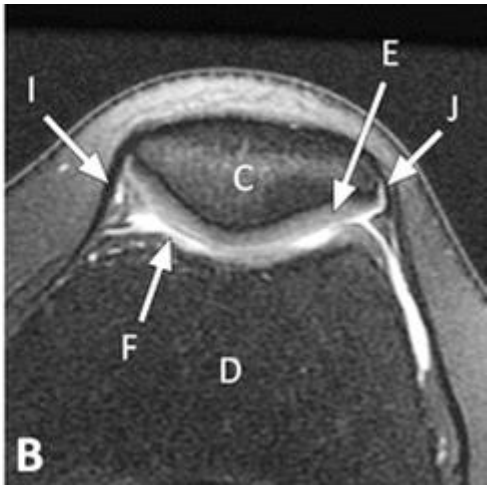
District / S	Total numl	Total numl	Number of	Number of	Number of	Number of	% 1st Trim	% 1st Trim	% JSY regis
_Hisar	34723	33456	26258	21979	12312	11440	75.6	65.6	35.5
Adampur F	383	0	156	0	85	0	40.7	NA	22.2
AryaNagar	3243	3260	2452	2000	1041	998	75.6	61.3	32.1
Barwala	4960	4860	4115	3282	2065	1876	83	67.5	41.6
District HC	NA	NA	NA	NA	NA	NA	NA	NA	NA
Hansi Urba	1101	936	377	355	375	353	34.2	37.9	34.1
Hisar Urba	5411	4396	2659	1867	1675	1302	49.1	42.5	31
Mangali	3643	3548	2799	2592	1240	1197	76.8	73.1	34
Narnaund	2999	3022	2711	2268	1057	1017	90.4	75	35.2
Sisai	3336	3516	2854	2550	1299	1199	85.6	72.5	38.9
Siswal	3499	3660	3015	2735	1117	1119	86.2	74.7	31.9
Sorkhi	2929	2965	2405	2166	965	981	82.1	73.1	32.9
Uklana	3219	3293	2715	2164	1393	1398	84.3	65.7	43.3

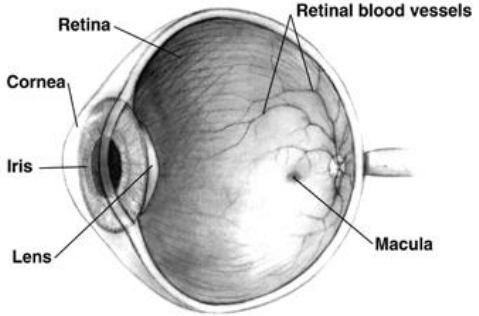
Spatial Data: Medical Images



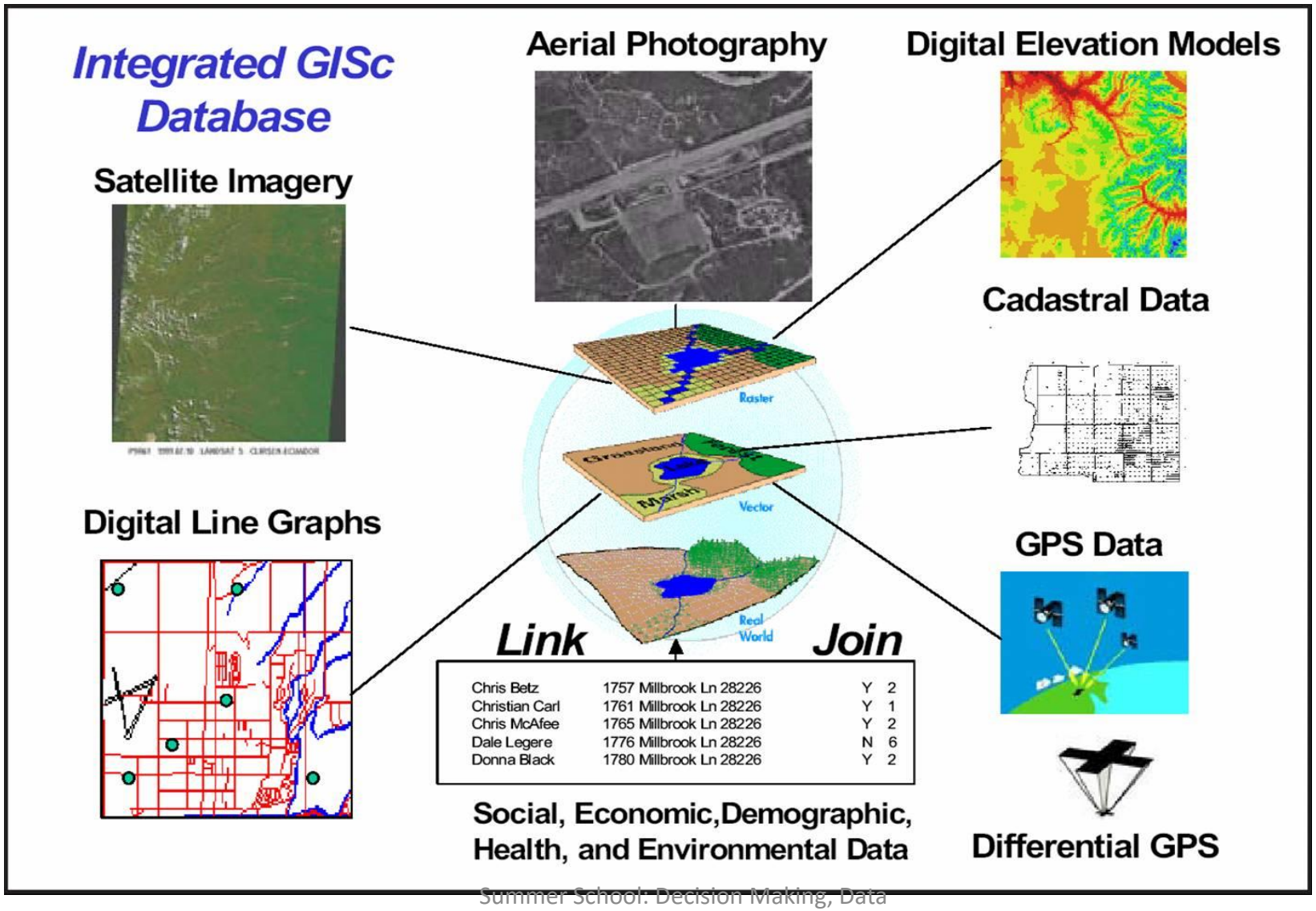








Geo-Spatial…


Integrated GISc Database
Aerial Photography
Digital Elevation Models
Satellite Imagery
Cadastral Data
Raster
Vector
Real World
Digital Line Graphs
GPS Data
Link
Join
Chris Betz 1757 Millbrook Ln 28226 Y 2
Christian Carl 1761 Millbrook Ln 28226 Y 1
Chris McAfee 1765 Millbrook Ln 28226 Y 2
Dale Legere 1776 Millbrook Ln 28226 N 6
Donna Black 1780 Millbrook Ln 28226 Y 2
Social, Economic,Demographic, Health, and Environmental Data
Differential GPS

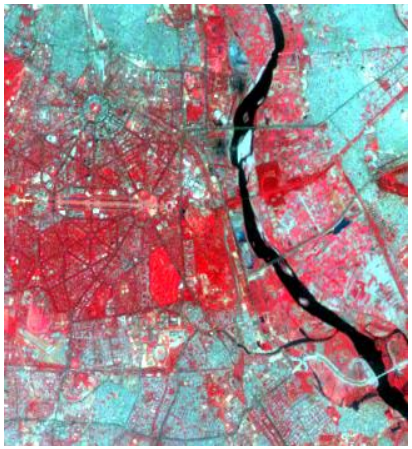
Knowledge-oriented Remote Sensing Image Analysis

Hemant Kumar Aggarwal, Sonajharia Minz, "*Change Detection Using Unsupervised Learning Algorithms for Delhi, India*," Asian Journal of Geoinformatics, Vol 13, No. 4, 12-15, 2013

Motivation

- Knowledge-oriented Change Detection
- Measure effectiveness of Machine Learning for Change Detection
- Explore potential for dimensionality reduction

Experimental Results

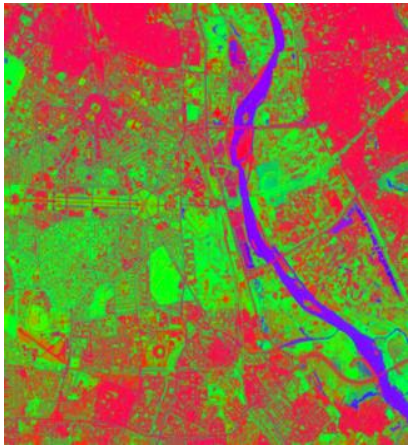


Original

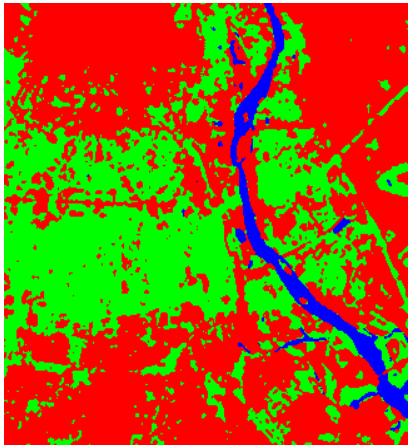
Water

Vegetation

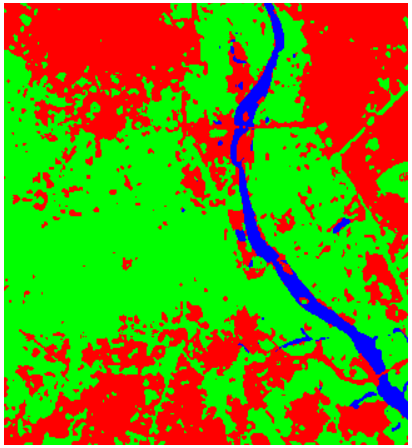
Urban



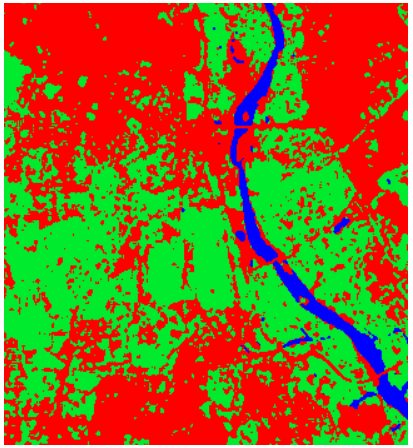
Features



K-means



Fuzzy C mean



EM

Percentage of Pixels per Class

Year	Kmeans	EM	FCM	Kmeans	EM	FCM	Kmeans	EM	FCM
1998	52.8	76.6	45.8	41.929	15.81	49.23	5.265	7.58	4.97
1999	35.16	55.43	43.25	59.19	41.22	53.97	5.648	3.35	2.78
2000	37.318	56	42.89	58.773	40.34	53.6	3.909	3.65	3.51
2001	36.016	38.72	51.85	59.881	55.78	44.45	4.104	5.49	3.69
2002	34.283	64.61	63.25	62.282	32.95	33.68	3.435	2.43	3.06
2009	37.867	35.025	38.02	58.837	61.19	58.79	3.296	3.78	3.18
2010	35.717	52.87	40.39	58.438	41.57	54.3	5.848	5.56	5.3
2011	34.662	58.54	36.16	61.904	37.11	60.56	3.434	4.35	3.27

Total Percentage Change

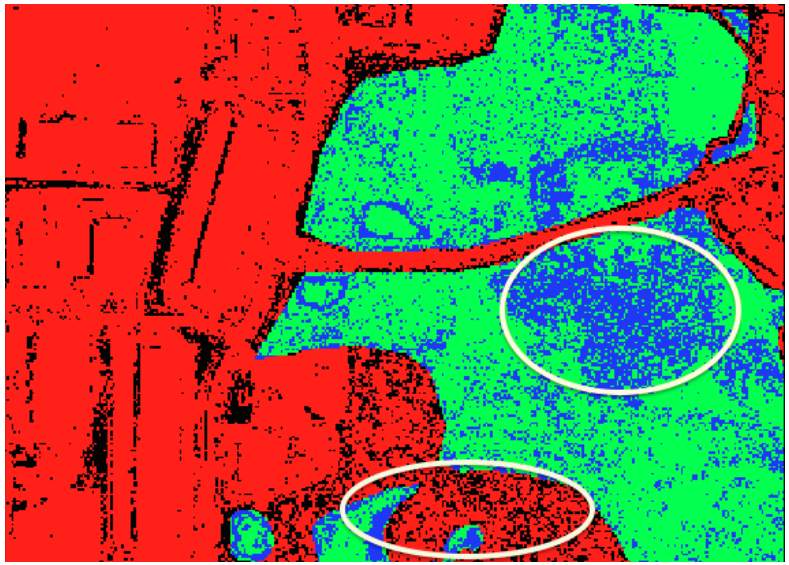
Class / Algorithm	Water	Built-up	Vegetation
K-means	-0.26	2.85	-2.59
EM	-0.46	3.04	-2.58
FCM	-0.24	1.62	-1.38

Conclusion

- Partitioning based methods are more effective than probabilistic and fuzzy.
- Decrease in vegetated area and increase in urban area.
- Dimensionality Reduction by 50%

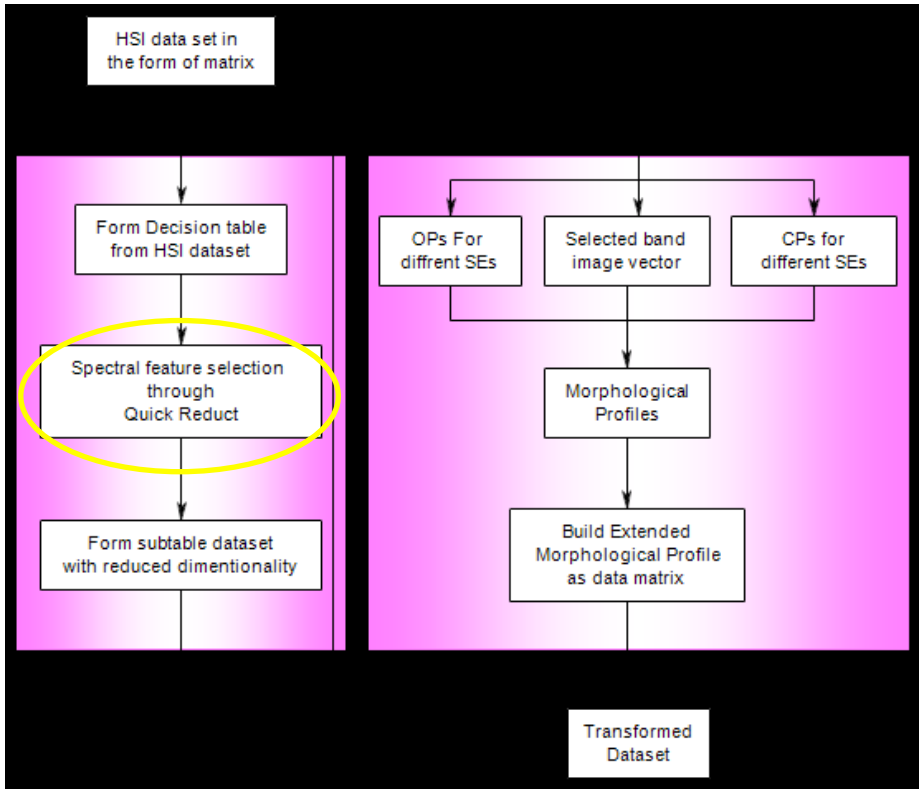
Future Work

- Environmental Footprint and More – Spatial Footprint Change Discovery

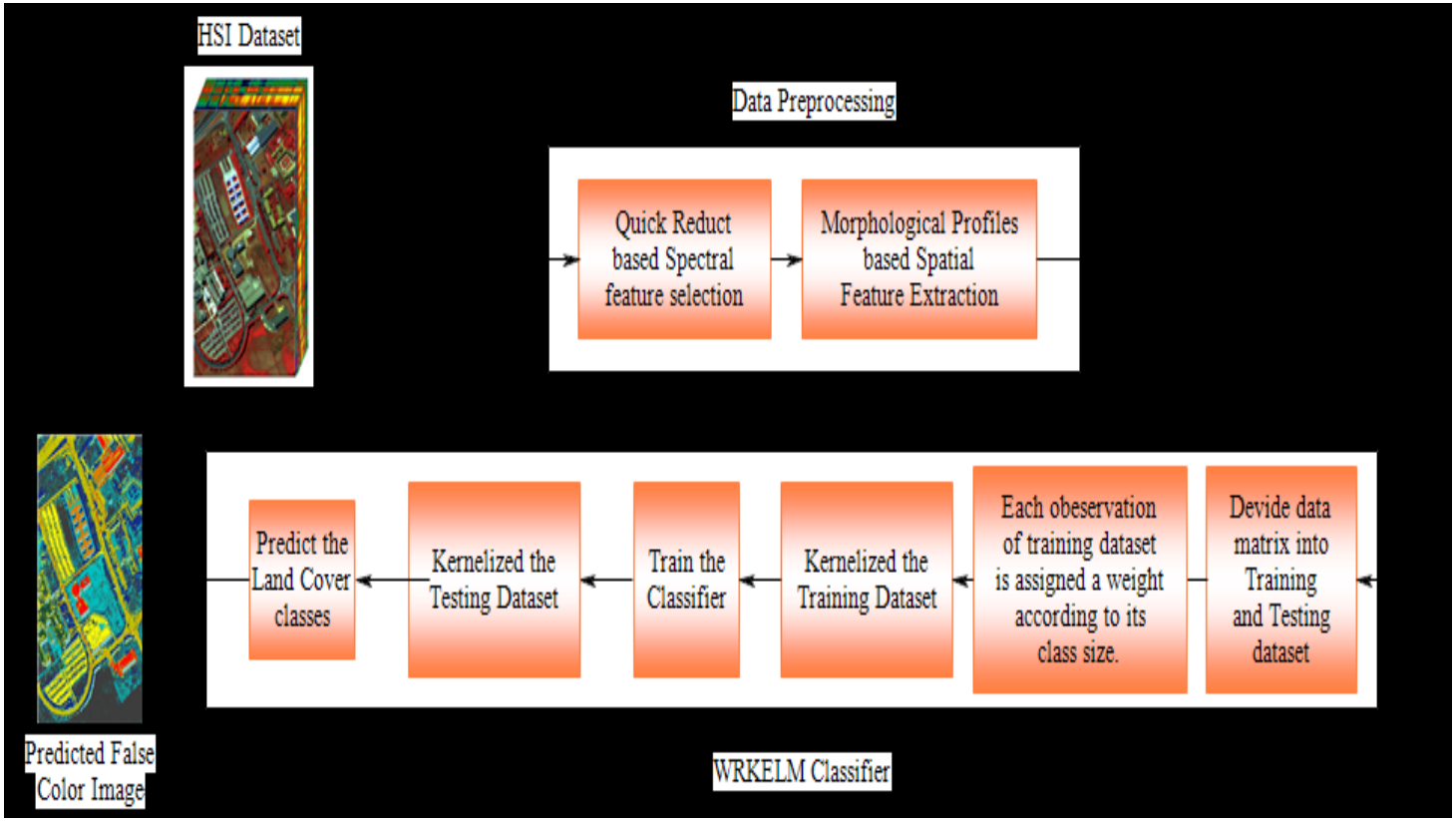


Rough Set based classification of Hyperspectral Data (Master's Dissertation: Ankit Malvia)

Framework


HSI data set in the form of matrix
Form Decision table from HSI dataset
Spectral feature selection through Quick Reduct
Form subtable dataset with reduced dimentionality
OPs For diffrent SEs
Selected band image vector
CPs for different SEs
Morphological Profiles
Build Extended Morphological Profile as data matrix
Transformed Dataset

Results: Dimension Reduction



S.N.	Datasets	Elapsed time	Total Bands	**No. Selected bands**	**Selected Features**
1.	Indian Pines	1156.42	200	**4**	**B1, B22, B43, B84**
2.	Pavia University scene	72165.33	103	**4**	**B1, B2, B87, B103**

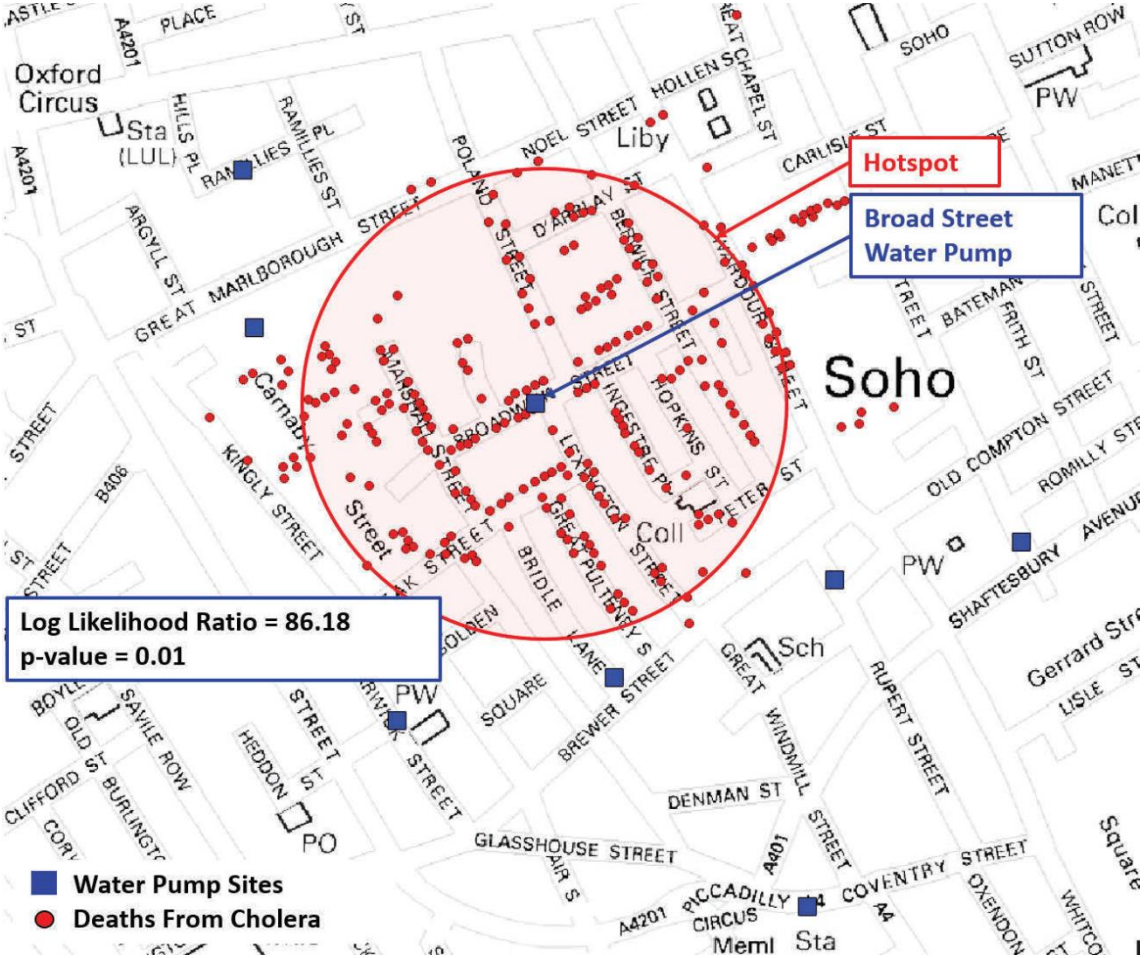
Spatial Computing

Challenges pertaining to Data Characteristics

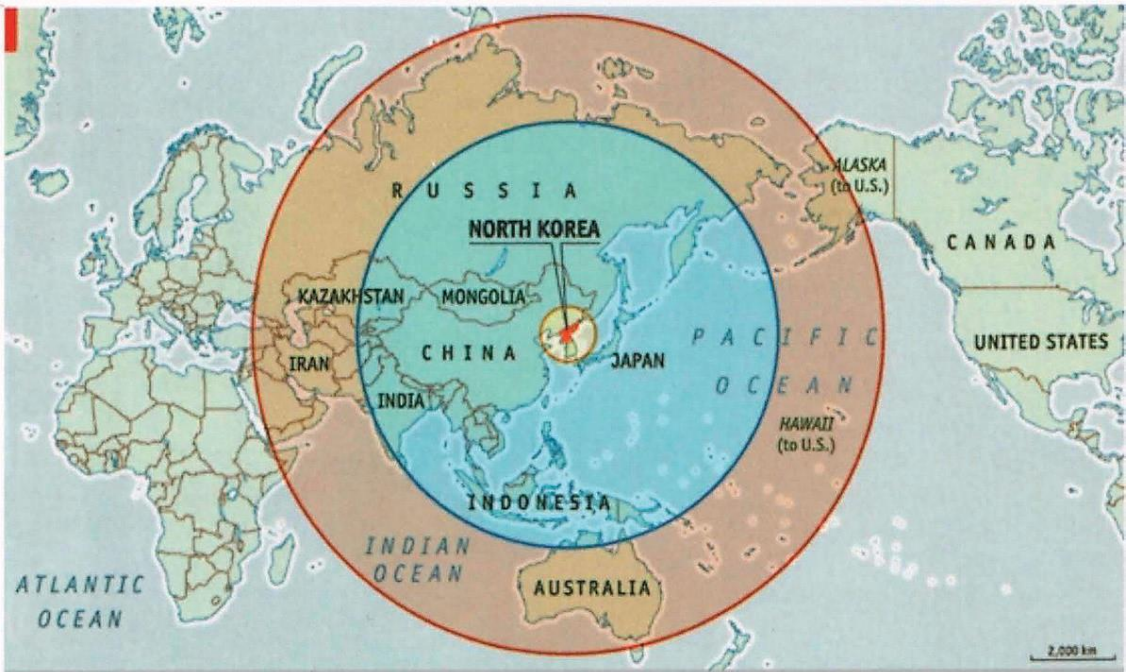
- Not iid – Spatial Auto correlation (Tobler's First Law of Geography – nearby objects are related to each other more than the objects at a distance)
- Class imbalance problem (challenge to Statistical and probabilistic methods)
- Classification – very small labelled data

Spatial Computing: Patterns

Spatial Patterns: Hotspot



Spatial Computing: Patterns

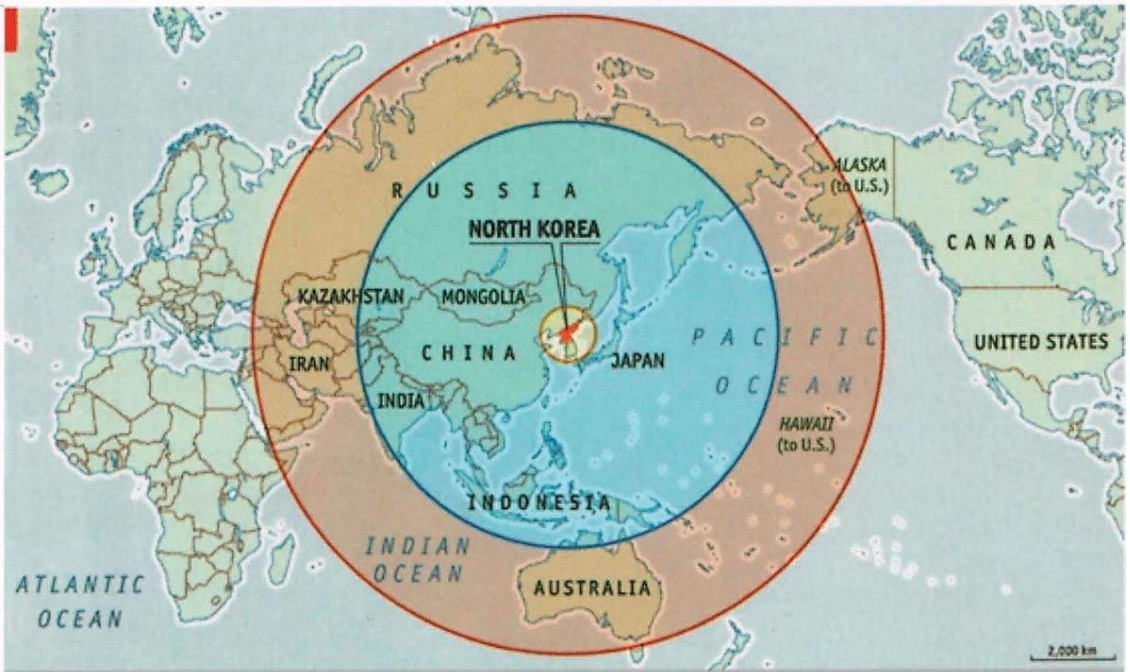


North Korea's missile threat

Type	Maximum range	Payload	Status
Nodong	1,300 km (810 miles)	700 kg (1,550 pounds)	Currently deployed
Taepodong-1	Up to 10,000 km	Several hundred kg	Test failed 1998, not yet operational
Taepodong-2	10,000-15,000 km	Several hundred kg	Not yet tested

Source: Task Force for US Korea Policy, Centre for International Policy

Patterns based on 2D plane



North Korea's missile threat

Type	Maximum range	Payload	Status
Nodong	1,300 km (810 miles)	700 kg (1,550 pounds)	Currently deployed
Taepodong-1	Up to 10,000 km	Several hundred kg	Test failed 1998, not yet operational
Taepodong-2	10,000-15,000 km	Several hundred kg	Not yet tested

Source: Task Force for US Korea Policy, Centre for International Policy

Spatial Patterns: Spherical Earth

North Korea's missile threat

Type	Maximum range	Payload	Status
Nodong	1,300 km (810 miles)	700 kg (1,550 pounds)	Currently deployed
Taepodong-1	Up to 10,000 km	Several hundred kg	Test failed 1998, not yet operational
Taepodong-2	10,000–15,000 km	Several hundred kg	Not yet tested

Source: Task Force for US Korea Policy, Centre for International Policy

Spatial Computing: Transformative Technology

- GPS
- Remote Sensing
- GIS
- Spatial Database Management Systems
- Spatial Statistics

Spatial Computing: Opportunities

- Short Term
  - Spatial Predictive Analysis
  - Geocollaborative Systems
  - Moving Spatial Computing
- Long Term
  - Fusion to Synergies
  - Sensors to Clouds
  - Spatial Cognitive first
  - Geoprivacy

Spatial Computing

- Shashi Shekhar: McKnight Distinguished University Professor, Department of Computer Science at the University of Minnesota, MN, USA
- https://vimeo.com/148128607
- http://cacm.acm.org/videas/sptial-computing

Some Important References

1. Komorowski, J., Pawlak, Z., Polkowski, L., Skowron, A., Rough Sets: A tutorial, 1997
2. Zadeh, L.A., Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets and Systems, 90, 1997
3. Zadeh, L.A., Granular Computing and Rough Set theory, LNAI 4585, 2007
4. Grzymala, J.B., Introduction to Rough Set Theory and Application, ppt
5. Jain, Rajni, Rough Set Theory based Decision Tree Induction for Data Mining, Ph.D. Thesis, 2005
6. Singh, Girish, Rough Set Theory for Data Mining, Ph.D. Thesis, 2007
7. Jain Rajni and Sonajharia Minz. 2008. "Drawing Conclusions from Forest Cover Type Data - The Hybridized Rough Set Model", *Journal of the Indian Society of Agricultural Statistics*, 62(1):75-84
8. Girish Kumar Singh and Sonajharia Minz. 2007. Discretization Using Clustering and Rough Set Theory. In: *Proceedings of International Conference on Computing: Theory and Applications (ICCTA'07).* ieeecomputersociety.org/10.1109/ICCTA.2007.51
9. Mushrif, M.M., Ray, A.K., A-IFS Histon Based Multithresholding Algorithm for Color Image Segmentation, IEEE Signal Processing Letter, Vol 16, No. 3, 2009

- Jain, Rajni and Sonajharia Minz. 2007. Intelligent data analysis for identifying rich: The rough set way. In: *Proceedings of 2nd National Conference on Methods and Models in Computing (NCM2C 2007)*, Eds: S. Minz and D.K. Lobiyal, JNU, New Delhi, Allied Publishers Pvt. Ltd., pp. 117-130.
- Minz, S. and Rajni Jain. 2005. Refining decision tree classifiers using rough set tools, *International Journal of Hybrid Intelligent System* 2(2):133-148.
- Rajni Jain and Sonajharia Minz. 2005. Dynamic RDT model for data mining, Proceedings of 2nd Indian International Conference on Artificial Intelligence (IICAI-05), Pune, India.
- Rajni Jain and Sonajharia Minz. 2005. Dynamic RDT model for mining rules from real data, *Journal of the Indian Society of Agricultural Statistics*, 59(2).
- Sonajharia Minz, Rajni Jain. 2003. Rough set-based Decision Tree model for Classification, Proceedings of 5th International Conference on Data Warehousing and Knowledge Discovery, DaWaK 2003 Prague, Czech Republic, *LNCS 2737*, 172-181. Rajni Jain, Sonajharia Minz. 2003. Classifying Mushrooms in the Hybridized Rough Sets Framework, Proceedings of *1st Indian International Conference on Artifical Intelligence (IICAI-03)*, Bhanu Prasad (Editor) 554-567.
- Sonajharia Minz, Rajni Jain. 2003. Hybridizing Rough set framework for Classification: An Experimental View, *Design and Application of Hybrid Intelligent Systems* A. Abraham et. al (Eds.), IOS Press, 631-640.
- Rajni Jain, Sonajharia Minz. 2003. Should decision trees be learned using rough sets?, Proceedings *of 1st Indian International Conference on Artificial Intelligence (IICAI-03)* , Bhanu Prasad (Editor)1466-1479.

Thank YOU