Soft Computing, Machine Intelligence and Granular Data Mining: Data to Knowledge

Sankar K. Pal
Indian Statistical Institute
Calcutta

http://www.isical.ac.in/~sankar
Contents

- What is Soft Computing?
- Pattern Recognition and Machine Intelligence
  - Relevance of Soft Computing Tools
  - Data Mining from PR point of view
- Fuzzy Sets and Flexibility
- ANN and GAs: Features
- Rough Sets and Information Granules
  - Example: Case mining
  - Other applications
Integrations of SC Tools: Challenges
- Rough-neural computing
- Neural-rough-fuzzy computing
- Rough-fuzzy computing
  - Generalized rough sets and entropy
- Examples
  - Object extraction in image/video

Challenging issues

Relevance to Big Data

Conclusions
SOFT COMPUTING (L. A. Zadeh)

Aim:

• To exploit the tolerance for imprecision uncertainty, approximate reasoning and partial truth to achieve **tractability, robustness, low solution cost, and close resemblance with human like decision making**

• To find an approximate solution to an imprecisely/precisely formulated problem.
Parking a Car

Generally, a car can be parked rather easily because the final position of the car is not specified exactly. If it were specified to within, say, a fraction of a millimeter and a few seconds of arc, it would take hours or days of maneuvering and precise measurements of distance and angular position to solve the problem.

⇒ High precision carries a high cost
The challenge is to exploit the tolerance for imprecision by devising methods of computation which lead to an acceptable solution at low cost. This, in essence, is the guiding principle of soft computing.
Soft Computing is a collection of methodologies (working synergistically, not competitively) which, in one form or another, reflect its guiding principle: **Exploit** the tolerance for imprecision, uncertainty, approximate reasoning and partial truth to achieve **Tractability, Robustness, and close resemblance with human like decision making.**

Foundation for the conception and design of high MIQ (Machine IQ) systems.
• At this junction, the principal constituents of soft computing are Fuzzy Logic FL, Neurocomputing NC, Genetic Algorithms GA, and Rough Sets RS.

• Within Soft Computing FL, NC, GA, RS are Complementary rather than Competitive.
Role of

**FL** : the algorithms for dealing with imprecision and uncertainty

**NC** : the machinery for learning and curve fitting

**GA** : the algorithms for search and optimization

RS → handling uncertainty arising from the granularity in the domain of discourse
Machine Intelligence: A core concept for grouping various advanced technologies with Pattern Recognition and Learning

IAS are physical embodiments of Machine Intelligence
Pattern Recognition System (PRS)

Measurement → Feature → Decision
Space       Space       Space

– Uncertainties arise from deficiencies of information available from a situation

– Deficiencies may result from incomplete, imprecise, ill-defined, not fully reliable, vague, contradictory information in various stages of a PRS
$M$ : Height, Weight, Complexion, Diet….

$F$: 

$D$ : Straight Line

$D \Rightarrow \text{Classifier Design}$
Clustering

Sex-wise

Father

Mother

Son

Daughter

Blood group wise

Age-wise
Tasks & Challenges

- **Classification**: Sampled data are given about the pattern space and the challenge is to estimate the unknown regions of the pattern space based on the sampled data (incomplete information) → Abstraction + Generalization

- **Clustering**: Entire data is given and the challenge is to partition it into meaningful regions. Number of regions may be known or unknown
Image Classification $\rightarrow$ Pixel Classification $\rightarrow$ Supervised
Image Segmentation $\rightarrow$ Pixel clustering $\rightarrow$ Unsupervised
Pattern Recognition and Machine Learning principles applied to a very large (both in size and dimension) heterogeneous database
≡ Data Mining

Data Mining + Knowledge Interpretation
≡ Knowledge Discovery

Process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data
Fuzzy Sets and Flexibility
FUZZY SETS

Classical set $\mu \in \{0,1\}$ Hard

Fuzzy set $\mu \in [0,1]$ Soft

$A = \{(\mu_A(x), x) : \text{for all } x \in X\}$

$\mu_A(x)$: degree of belonging of $x$ to $A$ or degree of possessing some imprecise property represented by $A$

Example: tall man, long street, large number, sharp corner, very young, etc.

- Fuzzy set is a Generalization of classical set theory

$\implies$ Greater flexibility in capturing faithfully various aspects of incompleteness or imperfection in a situation.
Fuzzy Sets are nothing but Membership Functions

Membership Function: Context Dependent
Flexibility of fuzzy set theory is associated with the Concept of $\mu$.

- $\mu$: A measure of compatibility of an object with the concept represented by fuzzy set.
- $\mu_{\text{TALL}} = 0.3$ means Compatibility of some one with the set ``TALL'' NOT the prob. that some one is TALL
  i.e., 0.3 is the extent to which the concept ``TALL'' must be stretched to fit him
- As $\mu$ Amount of Stretching Concept

- FUZZINESS IS ANALOGOUS TO ELASTICITY
Concept of Flexibility & Uncertainty Analysis
Relevance of Fuzzy Sets in PR

- Representing linguistically phrased input features for processing
  - Representing multi-class membership of ambiguous patterns
  - Generating rules & inferences in linguistic form
  - Extracting ill-defined image regions, primitives, properties and describing relations among them as fuzzy subsets
Combined choice

Correct       Wrong

First choice       Second choice

Conventional two states

Proposed four states

Combined choice

Wrong
Samples under combined choice, second choice can be corrected at higher level under the control of a supervisory programme.

Example: Satellite Imagery Analysis

Roads shaded with trees - Roads nearby water body or vegetation.

- INFORMATION on a pixel is not only from the concrete, but also from vegetation or water body.
- Decision on a pixel should indicate its multi-class membership with certainty values.
- Linking of Broken Roads can be guided with second / combined choice for their detection.
Calcutta (SPOT, Infrared) Enhanced Image
Second Combined First Choice

Pure Water Class
Linear Structures Segmented Image
Artificial Neural Networks (ANNs): Learning and Adaptation
Major Characteristics of ANN

- Adaptability to new data/environment
- Robustness/ Ruggedness to failure of components
- Speed via massive parallelism
- Optimality w.r.t error

- Machinery for learning (abstraction and generalization) and curve fitting
ANNs provide Natural Classifiers having

- *Resistance* to Noise,
- *Tolerance* to Distorted Patterns /Images (Ability to Generalize)
- *Superior Ability* to Recognize Overlapping Pattern Classes or Classes with Highly Nonlinear Boundaries or Partially Occluded or Degraded Images

- *Potential* for Parallel Processing

- Non parametric
Genetic Algorithms (GAs): Search and Optimization
Why GAs in PR?

- Methods developed for Pattern Recognition and Image Processing are usually problem dependent.
- Many tasks involved in analyzing/identifying a pattern need *Appropriate Parameter Selection* and *Efficient Search* in complex spaces to obtain Optimal Solutions.

Makes the processes
- *Computationally Intensive*
- Possibility of *Losing the Exact Solution*
GAs: Efficient, Adaptive and robust Search Processes, Producing near optimal solutions and have a large amount of Implicit Parallelism.

GAs are *appropriate* and *natural choice* for problems which need – Optimizing Computation Requirements, and Robust, Fast and Close Approximate Solutions.
Example of GA based Classification

- Automatic selection of no. of hyperplanes for approximating class boundaries for minimum miss-classification (VGA classifier)
- Chromosome (sexual) discrimination to reduce computation time (VGACD classifier)
- Robust Searching Ability (suitable when the search space is large)
SPOT Image of Calcutta in the Near Infra Red Band

(spatial resolution = 20m x 20m
wavelength = 0.79µm-0.89µm)

Scatter plot of the training set of SPOT image of Calcutta, containing seven classes.
Classified SPOT image of Calcutta (zooming the race course ‘R’ only) using (a) VGACD-Classifier, Hmax=15, final value of H=13, (b) VGA classifier, Hmax=15, final value of H=10, (c) Bayes maximum likelihood Classifier, (d) k-NN rule, k=1, (e) k-NN rule, k=3, (f) k-NN rule, k=\sqrt{n}.

Variation of the number of points misclassified by the best Chromosome with generations for VGACD classifier and VGA classifier
Rough Sets and Granular Computing
Rough Sets

\[ \Omega_B \subseteq U \]

Upper Approximation \( \bar{B}X \)

\[ x \]

Set \( X \)

Lower Approximation \( B X \)

\[ \left[ x \right]_B \text{(Granules)} \]

\[ \left[ x \right]_B = \text{set of all points belonging to the same granule as of the point } x \text{ in feature space } \Omega_B. \]

\[ \implies \left[ x \right]_B \text{ is the set of all points which are } indiscernible \text{ with point } x \text{ in terms of feature subset } B. \]
Approximations of the set $X \subseteq U$ w.r.t feature subset $B$

**B-lower:** $\underline{B}X = \{x \in U : [x]_B \subseteq X\}$  
Granules definitely belonging to $X$

**B-upper:** $\overline{B}X = \{x \in U : [x]_B \cap X \neq \emptyset\}$  
Granules definitely and possibly belonging to $X$

If $\underline{B}X = \overline{B}X$, $X$ is $B$-exact or $B$-definable

Otherwise it is Roughly definable

Rough Sets are Crisp Sets, but with rough description
Rough Sets

Two Important Characteristics

Uncertainty Handling
(Using lower & upper approximations)

Granular Computing
(Using information granules)
Cluster definition in terms of rough lower and upper approximations

- Lower and upper approximate regions could be crisp or fuzzy
In Real life problems –

- Sets and Granules can *either or both* be fuzzy \(\rightarrow\) Generalized Rough Sets

- Upper and Lower approx. regions could be crisp or fuzzy

- **Stronger** framework for uncertainty handling
- **Rough-fuzzy** computing : **New paradigm**
Before I describe the application of rough-fuzzy computing, let me explain the concept of $f$-information granules

Relevance of integration in SC paradigm
Concept of -

$d$- Information Granules using Rough Rules
• Rule provides crude description of the class using granule $\square$
Rule characterizing the granule can be viewed as the **Case or Prototype** representing the class/concept/region.

Elongated objects need multiple rules/granules.

*Unsupervised*: No. of granules is determined automatically.

Cases (prototypes) are granules, not sample points. **Case generation, NOT selection.**
Note:

- All the features may not appear in rules
  - Dimensionality reduction
- Depending on topology, granules of different classes may have different dimensions
  - Variable dimension reduction
- Less storage requirement
- Fast retrieval
- Suitable for mining data with large dimension and size
Example: IRIS data case generation

Three flowers: Setosa, Versicolor and Virginica
No of samples: 50 from each class
Features: sepal length, sepal width, petal length, petal width
Iris Followers: Setosa, Versicolor and Virginica

(a) Sepal L - Sep W
(b) Sepal L – Petal L
(c) Sepal L – Petal W
Iris Followers: Setosa, Versicolor & Virginica

(a) Petal L - Sepal W
(b) Petal W - Sepal W
(c) Petal W - Petal L
Iris Flowers: 4 features, 3 classes, 150 samples

Number of cases = 3 (for all methods)
**Information Granules**: A group of similar objects clubbed together by an indiscernibility relation

**Granular Computing**: Computation is performed using information granules and not the data points (objects)

- Information compression
- Computational gain
Applications of Rough Granules

- Case based reasoning (evident is sparse)
  - Case representation and indexing
- Prototype generation and class representation involving datasets large in dimension and size
  - Dimensionality reduction and Data mining
- Data compression and storing
- Clustering & Image segmentation (k selected autom)
- Knowledge encoding (NN structure formation)
- Granular information retrieval in heterogeneous media (e.g., text, hypertext, image) like WWW
Applications of Rough Granules

- Case Based Reasoning (evident is sparse)
- Prototype generation and class representation
- Clustering & Image segmentation (k selected autom)
- Case representation and indexing
- Knowledge encoding
- Dimensionality reduction
- Data compression and storing
- Granular information retrieval
Certain Issues

- Selection of granules and sizes
- Class dependent or independent
- Fuzzy granules
  - Fuzzy set over crisp granules
  - Crisp set over fuzzy granules
  - Fuzzy set over fuzzy granules

- Granular *fuzzy computing*
- *Fuzzy granular* computing

These issues would be addressed, in one form or others, in the following examples:
- Nature of granules - Role of granules - GFC or FGC - Superiority of R-F computing
• Individual Relevance of FL, ANN, GAs, RS to PR and mining Problems is Established adequately
Challenging Issues in Soft Computing Research: Judicious Integrations
In late eighties scientists thought –

Why NOT Integrations?

Fuzzy Logic + ANN
ANN + GA
Fuzzy Logic + ANN + GA
Fuzzy Logic + ANN + GA + Rough Set

Neuro-fuzzy hybridization is the most visible integration realized so far.
Why Fusion

**Fuzzy Set** theoretic models try to mimic human reasoning and the capability of handling uncertainty – (SW)

**Neural Network** models attempt to emulate architecture and information representation scheme of human brain – (HW)

NEURO-FUZZY Computing
(for More Intelligent System)
FUZZY SYSTEM

ANN used for learning and Adaptation — NFS

ANN

Fuzzy Sets used to Augment its Application domain — FNN
Recently -

Rough-fuzzy Computing:
A stronger Paradigm for Uncertainty Handling
Merits and Challenges

- GENERIC
- APPLICATION SPECIFIC
Certain Issues

- Selection of granules and sizes
- Class dependent or independent
- Fuzzy granules
  - Fuzzy set over crisp granules
  - Crisp set over fuzzy granules
  - Fuzzy set over fuzzy granules
- Granular fuzzy computing
- Fuzzy granular computing

These issues would be addressed, in one form or others, in the following examples:
- Nature of granules - Role of granules - GFC or FGC - Superiority of R-F computing
Rough-fuzzy Computing : Applications
Example:

Class-Dependent Rough-Fuzzy Granular Space and Classification

- Fuzzy granules in *modeling overlapping classes*

- Granules’ shapes are class *dependent*
- Rough sets are used on fuzzy granulated space for feature selection
- Effectiveness of Neighborhood rough sets is studied
- Fuzzy granules & Crisp computation → FGC
Fuzzy (f) granulation

Example:

Fuzzy granulation of features $F_1$ and $F_2$ characterizing granules for four overlapping classes

# of granules: $c^n$ vs. $3^n$ (l, m, h = 3)
Schematic diagram for pattern classification
Five classification models combining different granular feature spaces and feature selection methods

- Model 1: k-nearest neighbor (k-NN) classifier
- Model 2: CI fuzzy granulation + PaRS based feature selection + k-NN classifier
- Model 3: CI fuzzy granulation + NRS based feature selection + k-NN classifier
- Model 4: CD fuzzy granulation + PaRS based feature selection + k-NN classifier
- Model 5: CD fuzzy granulation + NRS based feature selection + k-NN classifier
Neighborhood Granule Generation for two overlapping classes

Two neighborhood granules centered at samples $x_1$ and $x_2$ in $F_1$-$F_2$ feature space. $\phi$ is the radius of the granules and $\Delta(x_i, x_j) \leq \phi$. Granules’ shape & size are determined by $p$ norm distance function ($\Delta$) and threshold $\phi$. 
Variation of classification accuracy with granule radius $\phi$ for three $p$-norm distances for model 5 and VOWEL data (Train set = 20%)

- Optimum $\phi = 0.45$
- Beyond 0.5, NRS based model can’t select relevant features to distinguish patterns, since possibility of possessing irrelevant/ contradictory feature information by granules increases
Multi-Spectral IRS Image of Calcutta

(Dim = 512x512, Spatial resolution = 36.25 m X 36.25 m, Wavelengths = 0.77-0.86\,\mu m, Major land covers = pure water, turbid water, concrete area, habitation, vegetation, open space)
Indices

Davies-Bouldin (DB) Index:

\[ DB = \frac{1}{c} \sum_{i=1}^{c} \max_{i \neq k} \left\{ \frac{S(v_i) + S(v_k)}{d(v_i, v_k)} \right\} \]

Dunn (D) Index:

\[ D = \min_{i} \left\{ \min_{i \neq k} \left\{ \frac{d(v_i, v_k)}{\max_{i} S(v_i)} \right\} \right\} \]

- S(v_i): Variance  \ d(.,.): Distance

- DB: for every i, it computes S & d values and {.} w.r.t. other k values, and then takes the max value of them; and then computes the average of c such values. (lower)

- D: for every i, it computes S & d values and {.} w.r.t. other k values, and then takes the min value of them; and then compute minimum of such c values. (higher)
$\beta$  

- $n$: total number of pixels in image  
- $\bar{x}$: mean gray value of the image  
- $x_i$: number of pixels in the $i$th ($I = 1, \ldots, c$) region obtained by a segmentation method.  
- $x_{ij}$: gray value of $j$th pixel ($j = 1, \ldots, n_i$) in region $i$  
- $\bar{x}_i$: the mean of $n_i$ gray values of $i$th region. Then  

$$
\beta = \frac{\frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2}{\frac{\sum_{i=1}^{c} n_i}{n} \times \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2} = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2}{\sum_{i=1}^{c} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}
$$

*Int. J Remote Sensing, 21(11), 2269-2300, 2000*
Multi-spectral IRS-1A image: Comparison of models
Four bands and partially labeled data (3439 out of 512x512) for six classes
Classified IRS-1A images (a) model 1 ($\beta = 6.86$, DB = 0.95) (b) model 5 ($\beta = 8.41$, DB = 0.73)
Zoomed region (bridge) of classified IRS-1A image with (a) model 1 (b) model 5
\( D(i) \) quantifies the dispersion of the misclassified patterns into different classes when the true class is \( i \).

Given an overlapping of a class with others, lower dispersion is desirable.
Dispersion score of R-F models for six classes of IRS-1A image

1: pure water (PW), 2: turbid water (TW), 3: concrete (CON), 4: habitation (HAB), 5: vegetation (VEG), 6: open spaces (OS)

Model 1: 1-NN
Model 2: CI FG + PaRS FS + 1-NN
Model 3: CI FG + NRS FS + 1-NN
Model 4: CD FG + PaRS FS + 1-NN
Model 5: CD FG + NRS FS + 1-NN
Computations time of R-F models with IRS-1A image

(512x512, 4-band image; # train samples 3439; p = 2, φ = 0.45; classes: PW, TW, concrete, habitation, vegetation, open space; MATLAB (matrix lab) environment in Pentium-IV with 3.19 GHz processor speed)

<table>
<thead>
<tr>
<th>Models</th>
<th>Computation time (Sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: 1-NN</td>
<td>305</td>
</tr>
<tr>
<td>2: CI FG+PaRS FS+1-NN</td>
<td>310</td>
</tr>
<tr>
<td>3: CI FG+NRS FS+1-NN</td>
<td>320</td>
</tr>
<tr>
<td>4: CD FG+PaRS FS+1-NN</td>
<td>330</td>
</tr>
<tr>
<td>5: CD FG+NRS FS+1-NN</td>
<td>340</td>
</tr>
</tbody>
</table>
Summary

- CD based f-granulation enables memb. functions to explore degree of belonging of features to different classes → better class label estimation

- NRS based feature selection (requires no discretization) facilitates to gather local information through neighbor granules for better class discrimination

- Classification performance of Model 5 with 10% training is even higher than models incorporating CI + (PaRS or NRS) with 50% training

  Significant when scarcity of training samples
So far FGC

Now Crisp granules & Fuzzy computation → GFC

So far Supervised

Now Unsupervised
Rough-Fuzzy Clustering & Uncertainty Analysis

Example:

Defining Class Exactness in terms of Granules
(Clustering – a basic module for data analysis and mining)

- Fuzzy sets enable handling of overlapping partitions
- Rough sets deal with vagueness and incompleteness in class definition
- Improved performance & faster than fuzzy clustering - GFC

Integrates the concepts of membership of fuzzy sets, and lower and upper approximations of rough sets into hard clustering.

While fuzzy sets enable handling of overlapping partitions, rough sets deal with vagueness and incompleteness in class definition.
only objects in boundary are fuzzified

- assign \( \mu_{ij} = 1 \) for objects in lower approx. region, while \( \mu_{ij} \) in \([0, 1]\) for those in boundary region

- assign higher weight for objects in lower approx region as compared to boundary region in computing centroids

- influence in computing centroids of own and other clusters (for lower – only on own centroid, for boundary – on all centroids)
Each cluster in rough-fuzzy clustering is represented by:

- a cluster prototype
- a crisp core (lower approximation)
- a fuzzy boundary
Rough-Fuzzy Clustering

- Provides a balanced mixture between
  - restrictive partition of hard clustering
  - descriptive partition of fuzzy clustering

- Faster than fuzzy clustering
- Better uncertainty handling/ performance
Brain MR Images (AMRI, Kolkata)

\[ c = 4 \]

Background, White matter, Gray matter, and Cerebrospinal fluid
Results on Brain MR Images

DB Index of Different C-Means

<table>
<thead>
<tr>
<th>Sample Images</th>
<th>DB Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HCM</td>
</tr>
<tr>
<td>2</td>
<td>FCM</td>
</tr>
<tr>
<td>3</td>
<td>RCM</td>
</tr>
<tr>
<td>4</td>
<td>RFCM(MBP)</td>
</tr>
<tr>
<td>5</td>
<td>RFCM</td>
</tr>
</tbody>
</table>

Results on Brain MR Images
(Pentium IV, 3.2 GHz, 1 MB cache, and 1 GB RAM)

Example: R-F c-means based segmentation

Scatter plot of two highest membs. of pixels for MRI Segmentation
(256x180 MRI image with 16 bit gray levels, # of pixels 46080; c = 4)

Average difference $\delta$ between two highest memberships of pixels:
$\delta = 0.145$ (low) at 1st iteration; $\delta = 0.652$ (high) at final iteration

LNCS Trans. on Rough Sets, 5390, 114-134, 2008
Rough-fuzzy -

- C-means clustering (numerical data)
- C-medoids clustering (string, relational data)
Application to -

Protein Sequence Analysis & Determination of Biobases (c-Medoids)
Application to -

Selection of Relevant Genes from Microarray Data using R-F Sets in Information Measure

- *Class independent* fuzzy granules for modeling Low, Medium & High from overlapping classes
- Gene selection → Maximization of relevance to decision attribute and minimization of redundancy with other genes
- Merits of FEPM
- Fuzzy granules & Crisp computation → FGC
- So far Image segmentation based on Pixel classification
- Now Image segmentation based on Gray level thresholding
- We define Generalized Rough Sets as stronger paradigm of uncertainty handling
Concept of -

Generalized Rough Sets

Incorporate fuzziness in set & granules of rough sets
Generalized Rough Sets

In practice, the Set and Granules, either or both, could be Fuzzy.

- Generalized Rough Set
- Stronger Paradigm for Uncertainty Handling
Generalized Rough Sets

When R is an equivalence relation

$X$ is a **crisp** set & Granules have **crisp** boundaries

$$RX = \{ u | u \in U : [u]_R \subseteq X \}$$

$$\overline{RX} = \{ u | u \in U : [u]_R \cap X \neq \emptyset \}$$

$[u]_R$ represents the granule that contains $u$.

The pair $<RX, \overline{RX}>$ is referred to as the rough set of $X$. 
When $R$ is an equivalence relation

$X$ is a **fuzzy** set & Granules have **crisp** boundaries

\[
\underline{RX} = \{(u, \inf_{z \in [u]_R} \mu_X(z)) \mid u \in U\}
\]

\[
\overline{RX} = \{(u, \sup_{z \in [u]_R} \mu_X(z)) \mid u \in U\}
\]

$\mu_X$ represents the membership function associated with $X$.

The pair $< \underline{RX}, \overline{RX} >$ is referred to as the rough-fuzzy set of $X$. 
R is an equivalence relation

$X$ is a **crisp** set & Granules have **fuzzy** boundaries

The pair $<RX, \overline{RX}>$ is referred to as the fuzzy rough set of $X$. 
R is an equivalence relation

$X$ is a **fuzzy** set & Granules have **fuzzy** boundaries

The pair $<RX, \bar{RX}>$ is referred to as the fuzzy rough-fuzzy set of $X$. 
Roughness Measure

\[ \rho_R(X) = 1 - \frac{|RX|}{|RX|} \]

- a measure of inexactness of \( X \)
- \( RX \) and \( \overline{RX} \) are the lower and upper approxs. of \( X \)
Entropy Measures using Roughness Values

Entropy measures based on roughness values of a set $X$ in $U$ and its complement $X^C$ are:

Measure using logarithmic gain function:

$$H^L_R (X) = -\frac{1}{2} \left[ \rho_R (X) \log_\beta \left( \frac{\rho_R (X)}{\beta} \right) + \rho_R (X^C) \log_\beta \left( \frac{\rho_R (X^C)}{\beta} \right) \right], \quad \beta \geq e$$

Measure using exponential gain function:

$$H^E_R (X) = \frac{1}{2} \left[ \rho_R (X) \beta^{(1-\rho_R (X))} + \rho_R (X^C) \beta^{(1-\rho_R (X^C))} \right], \quad 1 < \beta \leq e$$
Plots of entropy for different values of base $\beta$ and gain functions

$\rho_R(X)$

$\rho_R(X^C)$
Several Applications in Data Analysis

*Example: Image Analysis*

- **R-F entropy takes care of** - fuzzy boundaries of regions + rough resemblance between nearby gray levels + rough resemblance between nearby pixels (*i.e.*, *fuzziness* + *granulation*)

Nearby gray levels have limited discernibility

*Example: A region containing gray values separated by 6 gray levels.*
A grayscale image with sinusoidal gray value gradation. Boundaries cannot be defined exactly due to gray value gradation → Fuzziness

Example: A portion from the above image where the pixels in ‘white’ area belong uniquely to a region.

Nearby gray levels have limited discernibility

Example: A small region in the grayscale image containing gray values separated by 6 gray levels.

Granules
Entrophy based Grayness Ambiguity: Pixel Membership

- Set $X$ is fuzzy & Granules are crisp
- Pair $< RX, \overline{RX} >$ is referred to as the rough-fuzzy set of $X$

- x-axis: 0-N gray levels partitioned in crisp granules
- y-axis: $\mu$ values of pixels

- Fuzzy entropy: $\mu$ value of a pixel is *entirely dependent on its own gray value*
- Rough-fuzzy entropy: $\mu$ value is *dependent on the 1-d gray granule* to which it belongs
Example: Effect of fuzzy granules

- **Segmentation (and edge extraction):**
  Minimize $GA$ – w.r.t. crossover point of memb. function $\mu$ (assuming fuzzy set and fuzzy granules)
  - Membership of a pixel is *dependent on the 1-d gray granule* to which it belongs, and it is independent of its spatial location.
  - Results are compared to those of a fuzzy entropy with no concept of granule.
    - Membership of a pixel is *entirely dependent on its own gray value*, and it is also independent of its spatial location.

Difference is basically the effect of fuzzy granules.
Effect of granules
β-index for segmentation results on 45 images

*(window/granule size \( \omega = 6 \), Weber’s law)*

Significance of using the concept of \( f \)-granules is evident
So far we considered granules of equal size

Next, consider granules of unequal size
Example:

Formation of Unequal Granules and Spatial Segmentation

- Spatial Ambiguity (SA) measure
- Crisp set and crisp granules
- Granules formed by quad-tree decomposition
- Effect of granules of unequal size *vis-a-vis* fixed size
Example: Quad-tree decomposition and granule formation
Example Comparison

Original

Otsu’s thresholding

RE with 4x4 granule

RE with 6x6 granule

Rough-fuzzy with crisp set and 6x6 granule

Proposed methodology
Variation of $\beta$-Index over sequence ‘a’

- Homogeneous granules of unequal size reduce the formation of spurious segments → Reduce abrupt change of index-value over frames.
Video Tracking

- Spatial segmentation on each frame
- Temporal segmentation based on 3 previous frames
Relevance to BIG Data handling
Big-Data is

- High **volume** (scalable), high **velocity** (dynamic), high **variety** (heterogeneous) information
- Usually involves a collection of data sets so large and complex that it becomes difficult to process using conventional data analysis tools
- Requires exceptional technologies to efficiently process within *tolerable elapsed times*

**NEED** completely new forms of processing to enable enhanced decision making and knowledge discovery

- New approaches – challenges, techniques, tools & architectures to solve new problems
Dealing with big data
(Handling challenges lying with all Vs)

Demands a revolutionary change *both* in Research Methodologies and Tools
Example: PR (till 80’s) --> DM (since late Nineties)

- New approaches developed for different tasks of PR to handle DM problems (large data both in size and dimension)
- Example: Feature Selection - where instead of clustering samples in conventional PR, you cluster features themselves in DM
Dealing with big data: Challenges

- Challenges include - capture, preprocessing, storage, search, retrieval, analysis, and visualization
Dealing with big data: Tasks

- Tasks like:
  - Data size and feature space adaptation
  - Feature selection/ extraction in Big data
  - **Uncertainty modeling** in learning, sample selection, and classification/ clustering on Big data
  - **Granular computing** (a clump of objects…)
  - Distributed learning techniques in uncertain environment
  - Uncertainty in cloud computing

(Where SC methodologies can be used, in general)
In conclusion -

Without “Soft Computing” Machine Intelligence and Data Mining Research Remains Incomplete.
SK Pal and P Mitra, 
Pattern Recognition Algorithms for Data Mining, 
CRC/ Chapman & Hall, 
Florida, 2004
Thank You!!