Annals of the Indian National Academy of Engineering, Vol. XIII, pp. 25-34, April 2016

Prof. S.N. Mitra Memorial Award Lecture, INAE

50 Years of Fuzzy Sets: Data to Knowledge

Sankar K. Pal
Center for Soft Computing Research
Indian Statistical Institute
Kolkata 700108
http://www.isical.ac.in/~sankar

Abstract:

The theory of fuzzy sets (FS) was first explained in 1965 by Lotfi A. Zadeh, University of California, Berkely. The theory has been explored as a model of uncertainty analysis during the last fifty years by scientists over the globe for developing methodologies for decision-making problems of various kinds. The successful application areas and systems developed broadly include: fuzzy logic control based systems, fuzzy expert systems, and fuzzy pattern recognition and image processing systems. While the aims were to emulate and replace human operators and human expertise respectively in the first two categories, generalization and uncertainty handling were the objectives in the other one.

The talk addresses the characteristic features of fuzzy pattern recognition and image processing systems, mentioning the growth and evolution of the discipline, in the opinion of the author. It highlights the contributions made towards this from Indian Statistical Institute, Kolkata since early 1975. The talk starts with the relevance of FS to pattern recognition problems, and explains with examples the significance of multi-valued recognition systems and the notion of embedding. Data mining and knowledge discovery from pattern recognition perspectives is explained. Then follows the objective of synergistic integrations of FS with artificial neural networks, genetic algorithms and rough sets, among others, in soft computing for improved performance, computational intelligence and decision-making. Fuzzy-rough integration provides a stronger paradigm of modeling uncertainty arising both from vagueness and overlapping in data sets, and granularity in the domain of

discourse. Concepts of granular computing and generalized rough sets developed in the process for knowledge mining are explained. The significance of their application is demonstrated, as examples, in certain problems of image and video analysis, bio-informatics and social networks.

The talk concludes with future directions of research and relevance to Big data analysis.

Fuzzy sets:

The theory of FUZZY sets was introduced in 1965 by Lotfi A. Zadeh *(Inform. Control., 8, 338-353, 1965)* as a new way to represent vagueness in everyday life. They are a generalization of conventional set theory, one of the basic structures underlying computational mathematics and models.

A fuzzy set (A) in space of points X = {x} is a class of events with a continuum of grades of membership and is characterized by a membership function $\mu_A(x)$ which associates with each point in X a real number in the interval [0,1] with the value of $\mu_A(x)$ at x representing the grade of membership of x in A. Formally, a fuzzy set A with its finite number of supports x1, x2, , xn, is defined as

$$A = \{(\mu_A(xi), xi)\}, \qquad i = 1, 2, \ldots, n$$

The characteristic function $\mu_A(xi)$ denotes the degree to which an event xi may be a member of A, or possess some imprecise property as represented by A. As its value i, the grade of membership of xi in A becomes higher. In extreme cases when $\mu_A(x)$ is zero or unity, the fuzzy set A boils down to a crisp set. Typical examples of fuzzy set include tall man, large number, long street, very young, sharp corner, beautiful woman, which have no crisp or sharp boundaries.

Characteristics of fuzzy sets:

Fuzzy sets (FS) are nothing but membership functions, and membership functions are context dependent. That is, the function characterizing a set “Tall

Men” in India may not be the same as defined in case of Sweden, say, since Swedish are usually much taller than Indians.

FS is a *generalization* of classical set theory. That is, it has greater flexibility in capturing faithfully various aspects of incompleteness or imperfection in a situation. Flexibility is associated with the concept of $\mu(x)$. As the value of μ increases/ decreases, the amount of stretching required in fitting x with the imprecise concept, as represented by A, decreases/ increases. In other words, fuzzy sets are *elastic* and hard (crisp) sets are inelastic

It may be mentioned that $\mu(x)$ is different from $p(x)$, the probability of occurrence of x. While the former concerns with the compatibility (similarity) of x with an imprecise concept, the latter deals with the no. of occurrences of x.

Impact of FS on:

The following disciplines, among others, had impact of fuzzy set theory since 1965.

- Mathematics
- Modelling and Control
- Pattern Recognition (PR) & Image Processing (IP)
- Artificial Intelligence (AI)
- Data Analysis

Real world applications have been observed in various areas such as:

- Business and Finance Sectors
- Social and Behavioral Sciences
- Electronic and Robotic Systems (cars, domestic appliances)

The systems developed can be broadly classified as - fuzzy logic control based systems, fuzzy expert systems, and fuzzy pattern recognition and image processing systems. While the aim in the first two categories is to replace or emulate human operators and human expertise respectively, the objective in the other is to have generalization and uncertainty handling in decision-making.

One may note that Zadeh's first paper on fuzzy set theoretic pattern recognition. appeared in 1966 *(J. Math. Anal. & Appl. 13, 1-7, 1966),* while that for Control was published in 1973 *(IEEE Trans. Syst., Man and Cyberns., 3(1), 28-44, 1973).* It therefore appears that he had the concept (notion) of fuzzy classification of data while developing his theory.

Moreover, the techniques of pattern recognition (PR). & image processing (IP). interact with and support a large percentage of control systems (e.g., Mars rover control, camera tracking for repair & docking at space shuttles, fuzzy camcorders, traffic flow control). Applications of fuzzy PR. & Vision have *matured*, especially because of the commercial success of Japanese products based on fuzzy logic control.

The following discussions would be confined to issues concerning PR, IP, data mining and related areas only.

Relevance of fuzzy sets in pattern recognition:

The different tasks in PR and IP where fuzzy set theoretic approach has been found relevant and significant are as follows:

- 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

Linguistically phrased input means those described using the terms like small, medium, high etc.. For example, a patient usually describes his/her symptoms to a doctor in terms of those linguistic terms, viz, temperature is high, and pressure is low.

An ambiguous pattern (patterns lying in overlapping regions) should be characterized by multiclass memberships with values in [0, 1], unlike the conventional decision-making as characterized by {0, 1}, i.e., either belongs or does not belong. So, the decision on a pattern in c-class problem may be characterized by a membership vector with c components, where each component represents the degree of belonging to a particular class. These components and their statistics can be used to make decisions under different choices, like first choice, second choice and combined choice; thereby resulting in a multi-valued recognition system *(IEEE Trans. Syst., Man and Cyberns., 22, 607-620, 1992.).* Patterns under combined choice, second choice can be corrected at higher level under the control of a supervisory programme; thereby reducing the number of wrong decisions in conventional systems. For example, in defence applications, linking of bridge pixels and broken roads can be made under said supervisory scheme while processing remotely sensed imagery.

A gray tone image possesses ambiguity within pixels because of the possible multi-valued levels of brightness in the image. This indeterminacy (both in grayness & spatially) is due to inherent vagueness rather than randomness. Many basic concepts of image analysis (e.g., edge, corner, boundary, region, property, relation between regions) do not lend themselves well to precise definition. Therefore it is natural and appropriate to represent them as fuzzy subsets.

In summary, fuzzy PR and IP tasks involve a notion of embedding. We find a better solution to a crisp problem by looking in a large space at first, which has different (usually less) constraints and therefore allows the algorithm more freedom to avoid errors forced by commission to hard answers in intermediate stages.

Typical problems where application of fuzzy PR and IP has been successful include: Speech recognition, Medical image (MRI, X-rays), Remote sensing image (Defence applications), and Natural language processing.

It may be noted here that it is E. Ruspini (Stanford Research Institute), who first mentioned that clustering should be fuzzy, NOT crisp, i.e., patterns may have origin from more than one class *(Inform. Control, 15, 22-32, 1969)*. Subsequently, a new direction to fuzzy set theoretic cluster analysis was initiated by J.C. Dunn *(J. Cyberns., 3, 32-57, 1974; J. Cyberns., 4, 1-15, 1974)* and J.C. Bezdek *(J. Cyberns., 3, 58-73, 1974; J. Maths. Biol., 1, 57-71, 1974)* in their work on fuzzy ISODATA and fuzzy c-means algorithms. S.K. Pal and D. Dutta Majumder (ISI, Kolkata) published first IEEE paper in 1977 applying fuzzy sets in speech recognition. *(IEEE Trans. Syst., Man and Cyberns., 7(8), 625-629, 1977)*

Similarly, it is J.M.B. Prewitt (in *Picture Processing and Psychopictorics*, B. Lipkin and A. Rosenfeld (eds.), Academic Press, N.Y., 1970, pp. 75-149) who first mentioned that image segmentation should be fuzzy sub-sets of image, though not much work was done by her. Azriel Rosenfeld, father of image processing, and his group (UMD, College Park) contributed immensely in initiating fuzzy set theoretic image processing research. Their first article appeared in 1979 extending the concepts of digital picture geometry to fuzzy subsets *(Inform. Control, 40(1), 76-87, 1979).* Later on, his contribution in developing fuzzy geometry of image subsets (*Pattern Recog. Letters,* 2, 311-317, 1984) for soft decision-making is also noteworthy. S.K. Pal and R.A. King (when S.K. Pal joined Imperial College, London in 1979) are the other pioneers who published their first paper in 1980 *(Electronics Letters, 16(10), 376-378, 1980)* on fuzzy image enhancement. Subsequently, Pal and his group in ISI contributed significantly such as in defining fuzzy co-relation between membership functions *(Fuzzy Sets and Systems, 17, 23-38, 1985),* fuzzy primitives and grammars from images *(IEEE Trans. Syst. Man and Cyberns., 16(5), 657-667, 1986)*, fuzzy image entropy (*IEEE Trans. Syst, Man and Cyberns., 21(5), 1260-1270, 1991),* and geometry (*Fuzzy Sets and Systems, 48(1), 23-40, 1992)* for image segmentation and analysis, among others.

Around the same time when fuzzy pattern recognition algorithms were being developed, A. De Luca and S. Termini published their benchmark article on definition of a nonprobabilistic entropy in the setting of fuzzy set theory (*Inform. Control,* 20, 301-312, 1972). This has become a source of several later investigations on entropy and uncertainty measures with applications to PR and IP. The book by A. Kaufmann (*Introduction to the Theory of Fuzzy Subsets: Fundamental Theoretical Elements*, vol. 1. New York: Academic, 1975) was the only authored book available, and acted as valuable resource for beginners. The article on Zadeh's possibility theory based on fuzzy sets *(Fuzzy Sets and*

Systems,. 1, 3-28, 1978) is another land mark contribution towards approximate reasoning and decision-making in this regard.

Crisis in fuzzy logic research:

Research in fuzzy sets and logic got stuck little in mid '80s, as in many other areas and theories. One of the criticisms was concerned with the determination of membership functions

Meanwhile, Japanese products on FL control (e.g., Sendai, Japan subway system giving smooth ride) came to market vigorously with hundreds of patents. Artificial Neural Networks (ANN) re-appeared as machinery for learning and curve fitting. Theory of Genetic Algorithms (GAs) was introduced as a framework of parallel searching and optimization.

FL research flourished again at a higher gear in conjunction with ANN and GAs. Different funding agencies (both in India and abroad) came forward declaring it a thrust area. Various conferences/workshops held in conjunction with other paradigms and disciplines. The IEEE (USA) came forward by publishing different Transactions, and later on forming the Computational Intelligence Society. Other publishing houses (e.g., Springer, Elsevier) similarly brought several journals in the market in these and related domains. Special issues were also published by other engineering and applied science journals wherever applications of FL, ANN and GA were going on.

Different integrations:

In late eighties, scientists thought why not synergistic integrations among FL, ANN and GAs to overcome the limitations of individuals. For example, the searching characteristics of GAs can be used to determine the appropriate membership functions of fuzzy sets as well as in determining the optimal parameters of neural networks. Accordingly, FL+ANN, FL+GAs, ANN+GAs, FL+ANN+GA etc. based models were being framed with various applications. Among them, Neuro-fuzzy hybridization is the most visible integration realized so far. One of the reasons behind this N-F integration was as follows:

Fuzzy Set theoretic models try to mimic human reasoning and the capability of handling uncertainty Neural Network models attempt to emulate architecture and information representation scheme of human brain. So, if FL provides the software, ANN can provide the hardware. This led to the development of the Neuro-fuzzy computing paradigm in order to provide more intelligent systems.

There are two broad categories of integrations resulting in, namely Neuro-fuzzy systems and Fuzzy neural networks. Given a fuzzy system, ANN is used for efficient learning and adaptation of the rules in the first category. On the other hand, given a neural network, fuzzy sets are used at input, output and during training to augment its application domains. For example, in *IEEE Trans. Neural Networks,* 3, 683-697, 1992, the authors showed how the capability of generating nonlinear boundaries by multi-layer perceptron (MLP) and the uncertainty handling capabilities of fuzzy set theory can be combined to develop a network which can handle linguistic input in addition to those by conventional MLPs.

Soft Computing:

While different challenges of synergistic integrations between FL, ANN and GAs were being addressed with application specific merits, Zadeh defined the concept of Soft Computing consolidating them under one umbrela *(Comm. ACM, 37, 77-84, 1994).* The aim of soft computing is 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.* Mathematically, the objective is to find an approximate solution to an imprecisely/ precisely formulated problem. Since high precision carries high cost, the guiding principle of soft computing is to exploit the tolerance for imprecision by devising methods of computation which lead to an acceptable solution at low cost.

The roles of the constituting components are as follows:

FL : Algorithms for dealing with imprecision and uncertainty arising from overlapping concepts/ regions
ANN : Machinery for learning and curve fitting

GA : Algorithms for search and optimization

Later on, rough sets (RS), as explained by Pawlak in 1982 *(Int. J. Comp. Inf. Sci., 11, 341--356, 1982)* became another component of SC paradigm where its role is to provide, unlike FL, algorithms for handling uncertainty arising from granularity in the domain. Some of its characteristics would be described under the sections “Rough sets and cluster definition” and “Generalized rough sets”.

Within soft computing, FL, ANN, GA and RS work synergistically, not competitively. The SC framework provides *flexible information processing* in uncertain situations, and the foundation for the conception and design of high MIQ (Machine IQ) systems. It may be argued that it is *soft computing rather than hard computing* that should be viewed as the foundation for Artificial Intelligence.

An example of synergistic Integration of ANN, FL, GAs and RS can be found in *IEEE Trans. Neural Networks, 9, 1203-1216, 1998.* Here a layered network accepts input in terms of low, medium and high, and provides output in terms of class membership values. GA is used to replace the traditional gradient descent search techniques while finding the optimum network parameter values. In the process it provides a group of solutions (chromosomes) describing different parameter sets, and the possibility of getting stuck at local minima, as in gradient descent search algorithm, is greatly reduced. Rough sets are used to extract the domain knowledge from the training samples in the form of rough reducts. These reducts are encoded as initial network parameters so that the system starts learning from a better initial position; thereby reducing the training time drastically. Such a network, under split and merge modularity framework (*IEEE Trans. Knowledge Data Engg., 15(1), 14-25, 2003),* further enhances – the classification performance, training time and network compactness, and generates rules of higher accuracy, smaller size, and less confusion. Compact network means, all the links between nodes do not exist. That is, redundant (in the sense of decision-making) links get automatically discarded. This integrated framework became an established module for several future developments. Recently, various granular neural networks are developed for PR problems incorporating the concept of fuzzy-rough sets (in granular form) for network generation. Such an example case for unsupervised feature selection is available in *(Neural Networks, 48, 91-108, 2013.)*

Granulation – natural clustering:

In 1997, Zadeh explained the notion of granulation and its centrality in human reasoning and fuzzy logic (*Fuzzy Sets and Systems,* 90, 111–127, 1997). Granulation, meaning natural clustering, replaces a fine-grained universe by a coarse-grained one, more in line with human perception. Granules can be defined as a clump of indiscernible objects/points (i.e., similar objects, not discriminable by given attributes/relations).

Examples: Granules in

- Age: *very young, young, not so old,…*
- Direction: *slightly left, sharp right, ….*
- School: each class/section
- Image: *regions* of similar colors, gray values, e.g., max diff of 6 gray levels (Weber's law) cannot be noticed by naked eyes.

One may note that though Zadeh explained the notion of granulation in 1997, Pawlak's Rough set theory, which was defined long before in 1982, concerns with a granulated domain (crisp set defined over a crisp granulated domain).

Given an object or a region or concept over a granulated domain, whether labeled or unlabeled, rough set theory has the capability in extracting automatically some rule(s), called *Information granules*, that provide crude description of the said object/region/concept. These granules may be viewed as the cases or prototypes representing the regions. Elongated regions may need multiple such cases (rules). Since prototypes are rules, not sample points, these may be referred as case generation, not case selection.

In real life problems, all the dimensions of the feature space may not occur in these rules. That means, there is a possibility of dimensionality reduction. Again, depending on the topology of the feature space, granules (cases) of different classes may have different dimensions; thereby leading to the notion of *variable dimension reduction*. The cases thus obtained would have less storage requirement and fast retrieval; and are therefore suitable for mining data with large dimension and size.

Let us consider the Iris data which has three flowers (Setosa, Versicolor and Virginica) with four features (sepal length, sepal width, petal length, petal width) as an example. Here, two flowers are highly overlapped while the other is well separated. It has been observed (*IEEE Trans. Knowledge Data Engg., 16(3), 292, 2004*) that the representative cases generated by the aforesaid method need only 2.5 featutes on an average per case, whereas other well known methods need all the four features. This reduced set of features is also found to be superior to other methods in terms of classification accuracy, retrieval time and generation time.

Granular Computing (GrC)

GrC is a paradigm where computation is performed using *information granules* and not the data points (objects). Since information granules provide compressed information, GrC leads to computational gain, among other advantages, and is therefore suitable for mining large data sets.

Some of the proven applications of information granules are as follows:

- Case based reasoning (where evident is sparse)
- Class representation and indexing
- Clustering & image segmentation (initial classes selected automatically)
- Knowledge encoding in neural networks
- Dimensionality reduction
- Data compression and storing
- Granular information retrieval

It may be mentioned that it is rough set theory which has enriched the GrC research significantly. An excllent review on GrC is made by J.T. Yao, A.V. Vasilakos and W. Pedrycz (IEEE Trans, Cyberns., 43(6), 1977-1989, 2013.

Before we describe some applications of granular mining, we explain

- Rough set theoretic cluster definition, and
- Generalized rough sets (various fuzzy-rough sets) as a stronger paradigm of uncertainty handling

Rough sets and cluster definition:

In Pawlak's rough set, both the set and the granules are considered to be crisp. The set is defined over a granulated domain. The theory deals with the notion of belonging of a crisp granule to that crisp set. This is characterized roughly by the concepts of lower and upper approximations. Lower approximations denote the set of granules that do definitely belong to that set, while the upper approximations concern with those definitely as well as possible belonging. Accordingly, the crisp set defined over the granulated domain is characterized by that pair of crisp lower and upper approximations, and is termed as rough set. *Roughness of a set* is defined as: 1 - |lower|/ |upper|, where |.| denotes the cardinality of a set. It may be noted that, rough set, though the name is rough, is nothing but a crisp set with rough descriptions.

In the aforesaid rough set theoretic framework, a cluster or class in pattern recognition problems can be viewed in terms of lower and upper approximations. For example, samples in the central (core) region of a cluster/class have no ambiguity (doubt) in respect of their belonging to it; only ambiguity comes from those samples lying at its boundary regions. Therefore, the regions, where these two categories of samples have come from, may be referred to as lower and upper approximate regions respectively.

Generalized rough set and entropy:

In real life problems, the set and the granules that were considered in defining rough sets, either or both, can be fuzzy. Accordingly, lower and upper approximate regions would also be fuzzy, characterized by fuzzy membership functions. That is, every granule would have two membership values corresponding to lower and upper approximate regions signifying the degrees of belonging to the set, in case the set or the granules, either or both, are fuzzy. Various rough sets, so generated incorporating the notion of fuzzy sets, are called generalized rough sets which provide a stronger model of uncertainty handling (e.g., uncertainty due to both overlapping regions and granularity in domain) (*IEEE Trans. Syst, Man and Cyberns. Part B, 39(1), 117-128, 2009).*

Various entropy measures (using logarithmic and exponential gain functions) of generalized rough sets are defined in terms of set roughness to model the uncertainty arising from overlapping concepts and granularity. These can be used for analysing any kind of data, whether image data, web data, or protein sequence data, wherever uncertainty analysis is needed. Let us consider grey images as an example. Here rough-fuzzy (R-F) entropy takes care of the fuzzy boundaries (due to sinusoidal variation of intensity) of gray regions, and rough resemblance between nearby gray levels and rough resemblance between nearby pixels. Nearby grey levels have limited discernibility, i.e., rough resemblance. For example, a region containing gray values separated by 6 consecutive gray levels cannot be discriminated by naked eyes, and therefore can be considered as a granule. A pixel usually tends to attain a grey value close to those of its neighbors; thereby showing rough resemblance.

Various applications of these rough-fuzzy uncertainty measures and granular mining have been reported demonstrating their superiority over those defined using fuzzy sets or rough sets alone, and other conventional methods. Here we mention three such cases in the areas of image analysis, bioinformatics and online social networks, along with the role of granules.

Image segmentation and video tracking:

In the investigation reported in *IEEE Trans. Syst, Man and Cyberns. Part B, 39(1), 117-128, 2009*, the superiority of R-F entropy over fuzzy entropy for image segmentation is demonstrated both visually and quantitatively for different kinds of images. The threshold for segmentation was determined by minimizing the greyness ambiguity of the image plane. Here granules were formed to determine *roughly resembled* gray levels and pixels. Granules were of equal size of width (= 6 as per Weber’s law). In case of F-entropy, the membership of a level is determined uniquely by the membership function, irrespective of its location. On the other hand, for R-F entropy, the membership of a level depends on the granule to which it belongs. It has also been observed that the fuzziness in set has more effect than that in granules in segmentation.

Granules of un-equal size (which is natural) provide better segmentation than those of equal sizes. This is more apparent in case of video tracking based on spatial and temporal segmentation of each frame *(Applied Soft Computing, 3(9), 4001-4009, 2013)*. Here quad-tree spatial decomposition was made to produce homogeneous granules of unequal size. These natural granules of unequal size reduce the formation of spurious segments in frames, unlike the granules of equal size. In terms of index values (say, β, DB and Dunn indices), there would be an abrupt change (swing) in its value over frames if the frames are not properly segmented or produce spurious segments; thereby resulting in miss-tracking.

Gene selection from microarray data:

An important application of gene expression data in functional genomics is to classify samples according to their gene expression profiles. In most gene expression data, the number of training samples is very small compared to large number of genes involved in the experiments. Among the large amount of genes, only a small fraction is effective for performing a certain task. This led to gene selection problem, i.e., identifying a reduced set of most relevant genes.

Granules used here are class independent which model *low, medium & high* in the feature space for representing the overlapping classes. In this way, a fuzzy equivalence partition matrix (FEPM) of dimension (3xn) is generated; where n is the number of samples (objects) and 3 stands for *low, medium and high*. Using this fuzzy approximation space, various fuzzy *f*-information measured are defined. These include: Entropy (defined on fuzzy approximation spaces of a fuzzy attribute set *A*), Mutual information (defined between two fuzzy attribute sets *P* and *Q*), and other *f*-information measures, such as V & Chi-square information, between two fuzzy attribute sets *P* & *Q. (IEEE Trans. Knowledge & Data Engg., 22(6), 854-867, 2010*).

The principle of gene selection is based on *maximization* of relevance of a gene with respect to decision attribute (i.e., cancerous or normal) and *minimization* of redundancy with respect to other genes. Merits of FEPM based density approximation approach over those of discretization and Parzen-window for computing entropy, and mutual, V & Chi-square information have been adequately demonstrated for several data sets *(IEEE Trans, Syst., Man and Cyberns, Part B, 40(3), 741-752, 2010).* For example, for leukemia data with expression level of 7070 genes and 72 samples (47 acute lymphoblastic leukemia and 25 acute myeloid leukemia), the number of genes required to produce more that 98% classification accuracy is less than 10 by the proposed FEPM based method. Similarly for colon cancer with expression level of 2000 genes (with highest minimal intensity) of 40 tumor and 22 normal colon tissues, the proposed method selects only 7 genes to produce 90% accuracy. Out of ~18000 cDNA spots representing genes of relevance in immunology of 30 patients; 21 with rheumatoid arthritis (RA) and 9 with osteoarthritis, it needs only 3 genes to produce 100% accuracy. The corresponding figures of the required number of genes are higher for the conventional discretization and Parzen-window based approaches for computing entropy, and mutual, V & Chi-square information.

Community detection in social networks:

A social network is viewed as a collection of relations between social actors (nodes) and interactions between them. These actors are often indistinguishable in some problem solving, thereby justifying the formation of *granules* over them. Again, relations/ interactions between nodes & clusters of nodes do not lend themselves to precise definition, i.e., they have *fuzzy* boundaries. So, it is appropriate and natural that a social network is represented in the framework of *fuzzy granulation theory*.

Why community detection?

In society, one can find groups that naturally form, e.g., families, co-workers' circles, friendship circles, villages and towns. Similar to this, in an online social network, we can find virtual groups, which live on the web. Detecting these groups (communities) has practical significance. For example, in WWW this will help to optimize the internet infrastructure. In a purchase network this can boost the sell by recommending appropriate products. In computer network it will help to optimize the routing table creation.

Detecting these groups (communities) also help in identifying the special actors in the network. For example, central nodes of the clusters, or nodes in the boundary region (who act as a bridge between communities) are the special actors who play different important roles within the society.

FGSN model:

Recently, a fuzzy granular model for social network (FGSN) is developed *(Information Sciences, 314, 100-117, 2015)* using granules of various *hop* distances around each node. FGSN for undirected social networks is represented by a triple as S = (C, V, G) where V is a finite set of nodes of the network, C ⊆ V is a finite set of granule representatives and G is the finite set of all granules. The principle of community detection is as follows (*Pattern Recog. Letters*, *67, 145-152, 2015)*:

- Identify the dense fuzzy granules (whose granular degree exceeds a threshold, say θ) i.e., identify θ-*Cores*
- Merge them if they are nearby (search for θ-*Cores* belonging to same community i.e., find *Community reachable* θ-*Cores*)
- Form a meaningful community by discarding the weakly coupled granules (whose *Granular embeddedness* is less than a threshold, say $1/\tau$)

The communities C, thus detected, have fuzzy boundaries. Rough set theoretic view of fuzzy community C is made as follows:

- For nodes in lower approximate region reflecting their definitely belonging to C, assign membership value $\mu = 1$.
- For nodes in boundary (i.e., upper - lower) region reflecting their possibly belonging to C, assign μ-value in (0, 1).

Effectiveness of the aforesaid fuzzy-rough communities (FRC) over those obtained by three well known graph theoretic models is demonstrated on LFR

Benchmark data for various mixing parameters that signify different degrees of overlapping communities. The graph theoretic models which were compared are centrality based community detection method, modularity optimization method and k-clique percolation method. While the first two produce crisp communities, the other produces overlapping communities. The performance in community detection is evaluated in terms of normalized fuzzy mutual information (NFMI) index. NFMI measures the goodness of a community structure C, obtained by a method, given the actual one. FRC-FGSN model is found to be superior in detecting overlapping communities. Superiority is more prominent as the overlapping increases (*Pattern Recog. Letters*, *67, 145-152, 2015*).

Summary and future direction:

We described, in brief, the fuzzy set theoretic research from pattern recognition and machine intelligence perspective during the last fifty years in the world (including the contribution in forty years from ISI, Kolkata) since its inception in 1965. The evolution of the discipline over the years with different concepts, tasks and new technologies, driven by various applications areas as emerged time to time, is stated from the point of knowledge mining from data. This includes the crisis in FL research, synergistic integrations of fuzzy sets with other soft computing paradigms, rough-fuzzy computing and granular mining. What we discussed basically signifies the development of various efficient machine learning tools. These tools can be applied to any real life problem, although only a few example application cases are explained here.

Where are these researches leading to?

These researches have the significance, among others, to:

- *Computational Theory of Perceptions (CTP):* Here computation is performed based on perceptions, *not* measurements. Perceptions have ill-defined (fuzzy) boundaries and the attributes they can take are granules. This fuzzy (F)-granularity characteristics of CTP can be modeled using *fuzzy-rough* computing concept

In this context we mention the recent work on Z-numbers concerning perception granule in natural language processing. For a sentence like “The flower is beautiful”, perception granule is {subject, predicate, belief} = {description of flower, beautiful, how strongly one *Believes* to be beautiful}. (Belief could be “adverb”, or “adjective” or “adverbial phrase”. Belief = subjective probability that the flower is beautiful)

Z-numbers as explained recently by Zadeh *(Inform. Sci., 181(14),* 2923 – 2932, *2011)* provides precisiation of perception granules. On the other hand, Z*-numbers: Augmented Z-numbers *(Inform. Sci., 323(1), 143-178, 2015)* defined very recently represents machine–subjectivity in precisiation.

- *Natural computation:* Granulation is a process like self-reproduction, self-organization, functioning of brain, Darwinian evolution, group behavior, cell membranes and morphogenesis - that are abstracted from *natural phenomena f*-Granulation is inherent in human thinking & reasoning process, and plays an essential role in human cognition

- *Big Data research*: Any discussion today on data analysis remains incomplete without the mention of Big Data. The aforesaid research contributions have relevance to Big data handling and analytics (BDA) from uncertainty management and granular mining points of view. For example, on line social networks have all the characteristics of Big data such as large volume, velocity (dynamic) and varieties (complex). It may be mentioned here that the computational aspects and scalability issues have not been much addressed yet by the soft computing community.

All these constitute future research areas.

Note: Whatever is discussed during the talk and subsequently here in the article is entirely based on the opinion of the author. Since the duration of the talk was 30 minutes, I regret that the contributions and references of many other authors in the concerned domain could not be accommodated.

Acknowledgements

- Research fellows and younger colleagues/ collaborators both in India and abroad who contributed with me towards the development

- J.C. Bose Fellowship of the Govt. of India
- INAE Chair Professorship
- INAE selection committee for selecting me for 2015 Prof. SN Mitra Memorial Award

About the author

Sankar K. Pal received the PhD degrees from Calcutta University and Imperial College, London. He joined the Indian Statistical Institute in 1975 as a CSIR senior research fellow where he became a full professor in 1987, a distinguished scientist in 1998, and the Director in 2005. He is a J.C. Bose fellow of the Government of India and served as an INAE Chair Professor. He founded the Machine Intelligence Unit and the Center for Soft Computing Research at the Institute in Calcutta.

He worked at UC Berkeley and UMD, College Park, the NASA JSC, Houston, Texas, and the US Naval Research Lab, Washington DC. He has been a distinguished visitor of the IEEE Computer Society since 1987 and held several visiting positions in Italy, Poland, Hong Kong, and Australian Universities. He is a Life Fellow of the IEEE, and Fellow of TWAS, IAPR, IFSA, and all the four National Academies for science/engineering in India. He is a coauthor of 19 books and more than 400 research publications (with Google Scholar *h-index* close to 70, and 22,500 plus citations) in the areas of pattern recognition and machine learning, image processing, data mining, web intelligence, soft computing, bioinformatics, and cognitive machines. He is/was on the editorial boards of 20 journals including IEEE Transactions.

He received several national and international awards including the most coveted S.S. Bhatnagar Prize and Padma Shri in India, and Al-Khwarizmi International Award from the President of Iran.