Playing with Fuzziness and Ambiguity in Patterns – Challenges and Achievements

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

Background

I was born in Calcutta on September 13, 1950, the second of five siblings, to a very ordinary family of modest means. My father started as a clerk in Sree Saraswaty Press (a private concern at that time, now an undertaking of the Government of West Bengal) in Calcutta and retired, as far as I can remember, in 1978, as an Accountant earning five hundred rupees per month. So I was well aware of the hurdles I would have to face in completing school education, let alone higher studies. I am extremely fortunate in that my parents supported me wholeheartedly in my pursuit of a PhD degree, making great personal sacrifices in the process. I studied in four different schools. While the first three are in central Calcutta, the last one, Ariadaha Kalachand High School, is located in the northern suburbs of the city. I studied there for about four years and passed the Higher Secondary examination from there in 1966. Of those crucial four years, I spent three in my *mamar-bari* (house of maternal grandparents), as my parents' new residence, after relocating from central Calcutta, was rather inconveniently located for a safe daily commute. While the relocation from school to school and from an urban residence to a suburban one were primarily due to financial constraints, they did cause some amount of cultural, social and environmental shock.

Education

I gained admission to the B.Sc (Honours) programme in Physics at a Ramkrishna Mission college, named, Vivekananda Centenary College, Rahara which is located further north. The college had strict discipline. As there was no one in my family to provide proper academic guidance, I turned to Prof. Ananta Kumar Sengupta (father of one my class mates) who was then the Head of Applied Physics Dept., Calcutta University. I am extremely grateful to him for his valuable advice. In those days, the qualifying cutoff in marks required for admission to the B.Sc. (Honours) programme was the highest for Physics, among the pure science programmes. I had initially been interested in studying Chemistry, but changed my mind during admission. Some of my school teachers and relatives had discouraged/ cautioned me against joining the Physics programme thinking it to be too tough, but I took it on as a challenge.

After obtaining my undergraduate degree with honours, I had the option of joining either a two-year Masters' programme or one of the newly-introduced five-year B.Tech. cum M.Tech. programmes in Calcutta University. Despite the financial constraints, I opted for the latter, in Radiophysics in Electronics for better opportunity.

Initiating Research

After completing M.Tech. the programme, I joined the Indian Statistical Institute (ISI), Calcutta on March 01, 1975 as a CSIR Senior Research Fellow to work for a Ph.D. degree in the area of pattern recognition and man-machine communication by voice, a completely new discipline not related to any post graduate courses we had studied earlier. I had no idea about pattern recognition, nor were there any text books on this subject; only a few edited volumes, mostly by Prof. K.S. Fu, were available in our library or in the market. Apart from these, the D.Sc. thesis of G.S. Sebestyen, titled "Decision-making process in pattern recognition", published as a monograph by McMillan Press in 1962, was available in our library. From these, I started to pick up the basics of sequential pattern recognition using statistical approaches. One day my thesis advisor, Prof. D. Dutta Majumder, gave me a typewritten note, of about four pages, titled something like "Pattern classification with property sets", written by Prof. Ramesh Jain who had presented the concept in some seminar. There, I found a reference to Prof. Zadeh's famous paper, "Outline of a new approach to the analysis of complex systems and decision processes", which had appeared in the *IEEE Transactions on SMC* (vol. 3, pp. 28 – 44), in 1973. From these, I got some idea of multi-class belongingness of a pattern, which appealed to me very much as it seemed to be very natural for decision-making in real-life problems.

At that time my institute (ISI) library was not subscribing to almost any IEEE transaction, as Electrical Sciences or Computer Science was not considered to be a core subject of research there. However, I collected a copy from the Institute of Radiophysics and Electronics, Calcutta University, and then subsequently got the other two seminal papers of Zadeh which had appeared in *Information and Control* (1965) and *Information Sciences* (1967). Apart from these, I bought an edited volume from my research grant, which I consulted often. It was titled "Fuzzy sets and their applications to cognitive and decision processes" by Zadeh, Fu, Tanaka and Shimura, published by Academic Press in 1975. The Institute library subsequently acquired another book on fuzzy sets, by A. Kaufmann, titled "Introduction to theory of fuzzy subsets: Fundamental and theoretical elements", Academic Press, 1975, though only in the later part of my doctoral work. These were the only information sources on fuzzy sets to begin with, in the early part of my research career.

Visualizing Fuzziness

"Fuzzy sets are NOT *fuzzy!"*

In the following, I shall explain how I encountered fuzziness/ ambiguities in various problems of pattern recognition and image analysis, got motivated to handle them with fuzzy set theory, and subsequently continued to develop from time to time different hybrid technologies in the soft computing framework, as and when required to meet the then need. While doing so, I shall also describe the road map that I followed, acknowledging some persons concerned, and the different critical situations I faced.

A pattern recognition system (PRS) has basically three blocks, namely, measurement space, feature space and decision space. Uncertainties arise from deficiencies of information available in a situation. Deficiencies may result from incomplete, imprecise, ill-defined, not fully reliable, vague, contradictory information in various stages of a PRS. For example, vagueness can occur in the measurement space due to experimental error, limitation in instrument/ measurement to go for finer details, and availability of input in linguistic form. Sometimes, it may be convenient and appropriate to express the input feature value in interval form, or having one side of the interval unknown or even both sides fuzzy. In case of handwritten characters, for instance, vagueness comes from badness in writing, not from randomness. Accordingly, the resulting classes in the decision space may become non-convex, elongated, and overlapping, thereby making them intractable. These necessitate the design of classifiers with the capability of generating linear to highly nonlinear boundaries, or modeling overlapping class boundaries efficiently. In case of overlapping boundaries, it is natural and appropriate to make a multi-valued or fuzzy decision on an unknown pattern, i.e., a pattern has the possibility of belonging to more than a class with a graded membership. Depending on the distribution of membership values over different classes, the output decision on a pattern with respect to a class may therefore be *soft* and linguistically quantified as - “definitely belongs”, “definitely does not belong”, “combined choice”, ”second choice” etc. For a doubtful pattern it is always better to say “doubtful”, rather than misclassifying it. In that case the aforesaid multi-valued decisions have at least an opportunity to get the pattern correctly classified with some higher level information (e.g., syntactic, semantic), if available. This, in turn, dictates that a good classifier should be able to restrict the misclassified samples within less number of classes.

I was interested in working on speech and speaker recognition problems. Speech, being patterns of biological origin, their characteristics depend greatly on speakers’ health, sex, age, temperament, spirit and mind; thereby resulting in considerable amount of fuzziness in them and overlapping among the classes. For example, the same word uttered by a speaker at different times in a day may have different characteristic features. Accordingly, I started developing methodologies for nonparametric classification and recognition, and published my first IEEE paper – “Fuzzy sets and decision-making approaches in vowel and speaker recognition”, *IEEE Trans. SMC,* 7, pp. 625-629, 1977. Subsequently, I published on plosive recognition and self-supervised adaptive recognition systems, and submitted my PhD thesis to Calcutta University in 1978 titled "Some studies on pattern recognition and man-machine communication by voice with fuzzy set theoretic approach”. The foreign examiner of my thesis was Prof. K.S. Fu, the father of pattern recognition, Purdue University, USA. While appreciating the work as a pioneering contribution, he envisaged a future problem to study the sensitivity of different fuzzifiers and hedges, which were used in the distance function and similarity measure, on recognition performance. Prof. Zadeh mentioned in the Foreword of my 1992 IEEE Press book, co-edited with Jim Bezdek, that “S.K. Pal first applied fuzzy sets to the speech recognition problems in 1977”.

Meanwhile, I started realizing that the processing of gray images could be another good candidate area for fuzzy set theory application. Since it is gray, the basic concepts of image regions, segments, edges, skeletons and relations among them etc, do not lend themselves to precise definition. For example, a question like – “Where is the boundary?” – has no precise answer. Whatever hard decision that one may make for extracting those features/ primitives would always lead to an uncertainty. In other words, it is appropriate and also natural to consider the various tasks of processing of a gray image to be fuzzy, NOT hard, to manage the associated uncertainty in processing as well as in recognizing the content. Again, in an image recognition or vision system, once an uncertainty is caused in edge detection, segmentation, skeleton extraction etc. on account of the application of hard decisions (0 or 1) at the processing stage, it is likely to propagate further to the primitive extraction stage and finally affect the decision-making process in identifying the image contents. This further justifies the significance of fuzzy processing whereby the uncertainty can be minimized at the final stage of a vision system by retaining the gray information in the preceding stages as much as possible, and the ultimate output will not then be biased/ affected much by lower level decisions. One may note that gray information is very informative and expensive too; once they are made crisp by a threshold, the information is lost and can no way be retrieved. At the point of final decision-making at the highest level, one can always make them binary.

I was then looking for an opportunity to work in image processing. At that time, labs with complete software and hardware facilities for working in (gray) image processing were not readily available in many universities/ institutes, not even in the developed nations. Luckily, I got a Commonwealth Scholarship to study at Imperial College, London in 1979. (Though there was a possibility to work at Purdue University, USA, as a Post-doc Fellow with Prof. K.S. Fu, I chose to go to Imperial College for a 2nd PhD. To me, publishing quality research papers is the primary concern, PhD is a by-product.) Some digitized image data (in paper tape) was collected from Philips Research Laboratory, Redhill, Surrey. Then I started developing various algorithms for enhancement including image definition, edge detection, primitive extraction and image entropy measures using fuzzy sets, and publishing them in IEEE Transactions and Electronics Letters. Subsequently, I obtained another PhD in Electrical Engineering in early 1982 in the area of fuzzy image processing. In a gray image there are two types of ambiguities, namely, grayness ambiguity and spatial ambiguity. The former is concerned with whether a pixel can be considered to be black or white, and depends only on the gray value, whereas, the latter is concerned with both the gray level and location of pixels characterizing the geometry of image subsets. In the process of formulating the algorithms, Zadeh's contrast enhancement operator (INT), S & π membership functions, max & min operators, index of fuzziness, and entropy of fuzzy sets were used. With contrast enhancement of a fuzzy image around a fixed cross-over point, the difficulty in deciding whether a pixel is black or white reduces, and accordingly the values of its index of fuzziness and entropy decrease (*IEEE Trans. PAMI,* 4, pp. 204-208,1982). Similarly, given a set of fuzzified versions of an image, the one with minimum index of fuzziness or entropy gives the best segmented output for object extraction.

As an application, we choose the problem of identifying different stages of skeletal maturity (growth) with age from x-ray images of radius and ulna of wrist. The problem is significant from the point of determining the various stages of malnutrition of babies. We collected the image data from Prof. L.F. Turner, Institute of Sick Children, London. Here the shapes of radius and ulna at several stages of growth have overlapping character, i.e., they look alike. Accordingly, these were handled with fuzzy syntactic recognition approach, where both the primitives (e.g., vertical, horizontal & oblique lines, and curves) and the relations among them were considered to be fuzzy in developing the unambiguous grammars using production rules. Since the same set of production rules with different membership functions characterizes more than a class, the number of rules required is less, as compared to those using deterministic rules (*IEEE Trans. SMC*, 16, pp. 657-667, 1986.)

It may be mentioned here that my thesis advisor Dr. Robert A. King, Department of Electrical Engg., was basically an expert in the area of signal processing, and had not worked earlier on fuzzy set theory or image processing, till I joined him. However, he was convinced about my ideas and allowed me to work independently to develop the subject. One may further note that there was another pioneering group on fuzzy image processing led by Prof. Azriel Rosenfeld, Univ. of Maryland, College Park, father of image processing, working since late seventies, particularly in fuzzy geometry, connectedness and topology on image subspace, among others.

After returning to India in May 1983, I started developing, with my students, multi-valued recognition systems with linguistic input, fuzzy syntactic recognition methods, and various entropy measures and image segmentation algorithms, among others. We have defined correlation between fuzzy sets, and fuzzy operators using ordinary sets. Problems like estimating the entire class from a set of few sampled patterns, selected randomly, were dealt with fuzzy sets. In the area of image analysis, we have given various definitions of image entropy based on exponential gain function, and other quantitative indices for image processing tasks. The exponential gain function relies on the fact that a better measure of ignorance is $(1 - p_i)$ rather than $1/p_i$ (as used by Shannon), where p_i is the probability in receiving the i_{th} event *(IEEE Trans. SMC,* 21, pp.1260-1270, 1991). Accordingly, we have defined higher order fuzzy entropy, image entropy and hybrid entropy. As the order of image entropy increases, the validity of the segmented outputs, with respect to minimizing uncertainty, becomes more meaningful and valid. Hybrid entropy takes care of both probabilistic and fuzzy entropy and has significance in digital communication, particularly in noisy environment, where the concern is whether a bit is transmitted or not in a noisy channel and if its exceeds a threshold or not.

During 1986-1987, I visited the University of California, Berkeley and the University of Maryland, College Park as a Fulbright Fellow. That was the first time I met Prof. Lotfi A. Zadeh and Prof. Azriel Rosenfeld in person. Among the several characteristics of Lotfi, two features that appeared to be unusual and thus impressed me are – when I wanted to write a paper with him, Lotfi told me that he loves to work alone (showing his list of publications), and advised me not to put his name; and he never discussed fuzzy sets when we were together, whether in a car, or a restaurant or at his house.

My reminiscences and road map would remain incomplete, if I do not mention the criticism that I received often, like many other fuzzy researchers, from my colleagues when delivering lectures or seminars within my Institute and outside. The situation can be felt easily considering that I have been in an organization named, Indian Statistical Institute, surrounded by probabilists and statisticians. However, we have always viewed it as follows:

- Fuzzy set theoretic approach supplements the probabilistic approach and it is not a competitor, rather provides enrichment.
- We find a better solution to a crisp problem by looking at a larger space at first, which has different (usually less) constraints and therefore allows the algorithm more freedom to avoid errors by commission to hard answers in intermediate stages – *notion of embedding*.

Neuro-Fuzzy and Rough-Fuzzy Computing

In the late nineteen-eighties, I got interested in neural networks, and started developing with my students various neural and neuro-fuzzy models mainly for classification, clustering, rule generation, feature selection and connectionist knowledge based systems. The idea of synergistic integration was to enable ANNs to accept linguistic input (*low, medium, high, missing features*) in addition to numerical input; exploit the ANN characteristics like adaptivity, robustness, ruggedness, speed via massive parallelism, optimality and capability in generating highly nonlinear boundary; and uncertainty handling capability of fuzzy sets in the input, output and during training. This greatly enhances the application domain of ANNs. Our article - "Multi-layer perceptron, fuzzy sets and classification" (*IEEE Trans. NN,* 3, pp. 683-697, 1992) received the *Outstanding paper award from IEEE Neural Networks Council*. We have developed a series of generic models and demonstrated their applications to noisy/overlapping fingerprint identification, speech recognition, atmospheric science, image processing etc. Through integration, it has also been possible to make a layered network, which is usually used as supervised classifier, act as an unsupervised classifier using the fuzziness measures as error detectors (*IEEE Trans. FS,* 1, pp. 54-68, 1993). Efficient unsupervised feature selection models were also developed accordingly (*IEEE Trans. NN,* 11, pp. 366-376, 2000) along with some application specific models e.g., mixed category perception (*IEEE Trans. NN,* 6, pp. 1091-1108, 1995) and fuzzy clustering network for hardware realization (*IEEE Trans. NN,* 11, pp.1174-1177, 2000).

To enhance the computational intelligence characteristics of the said fuzzy networks, particularly for mining large data sets, we then started integrating the merits of rough sets and genetic algorithms into them. I had become interested in rough sets (RS) and genetic algorithms (GA) when I was visiting the NASA Johnson Space Center, Houston, TX during 1990-92 and 1994 as an NRC Senior Research Associate. I used to attend several seminars on rough sets organized in the Software Technology Branch, Information Technology Division. Two features of rough sets, namely, granular computing with information rules and uncertainty analysis with lower and upper approximations drew my attention. Since RS has the capability in extracting the domain knowledge, whether supervised or unsupervised, with reduced dimension in the form of information granules/rules, these can be encoded as initial network parameters for reducing its learning time significantly. Similarly, GA based learning, with chromosomes defined in terms of the network parameters and modified genetic parameters (*IEEE Trans. SMC-B,* 28, pp. 816-828, 1998), can replace the traditional gradient descent search technique which is slow and often gets stuck at local minima. The aforesaid synergistic integration of the four tools in the soft computing paradigm results in gain in terms of performance, computation time and compactness of the network, among others (*IEEE Trans. KDE,* 15, pp.14-25, 2003). So, it has wide application in mining data sets with large dimension and size, and in knowledge discovery. In this context, the convergence of GAs, which was a long standing problem, was also proved theoretically (*Int. J. Patt. Recog. & Arti. Intell.,* 10, pp.731-747, 1996).

Meanwhile I also realized that since both fuzzy sets and rough sets provide algorithms for different kinds of uncertainty, why not integrate them to have a much stronger paradigm for uncertainty handling than either of them. In 1997 I visited Prof. Andrzej Skowron, Warsaw University, Poland under an INDO-POLISH collaborative project; there I met Prof. Z. Pawlak, father of rough sets. Andrzej and I edited a volume - Rough-Fuzzy Hybridization: A New Trend in Decision Making, Springer, Singapore, 1999, which is the first of its kind.

One may note that Pawlak's rough set theory is based on the concept of crisp set and crisp granules, and provides a framework of handling uncertainty arising from granularity in the domain of discourse or limited discernibility of objects. However, in real life problems, one or both of them may be fuzzy. A gray image is such an example where the set (e.g., object region) can be fuzzy and the granules (e.g., pixel windows) may be overlapping. In order to model this, we have defined generalized rough sets (*IEEE Trans. SMC-B,* 39, pp. 117-128, 2009), where the set and granules could be crisp as well as fuzzy. Accordingly one could use "granular *fuzzy computing*" or "*fuzzy granular* computing" depending on the application.

For example, in an image nearby gray levels have limited discernibility, i.e., nearby gray levels *roughly resemble* each other and the values at nearby pixels have *rough resemblance*. Therefore, in the rough-fuzzy computing framework image ambiguity may be viewed as resulting from fuzzy boundaries of regions + rough resemblance between nearby gray levels + rough resemblance between nearby pixels. Accordingly, we defined generalized rough-fuzzy entropy based on lower and upper approximations. Its merits over fuzzy entropy and the significance of fuzzy granules have been demonstrated (*IEEE Trans. IP,* 18, pp. 879-888, 2009) as well as shown the superiority of unequal-size granules to equal size (*Applied Soft Computing,* 13, pp. 4001-4009, 2013).

Merits of rough sets and fuzzy sets have also been integrated judiciously in clustering problems where rough sets deal with vagueness and incompleteness in class definition, and fuzzy sets enable handling of overlapping partitions. Each cluster here is represented by a cluster prototype, a crisp core (lower approximation) and a fuzzy boundary. Membership values are unity for the objects in the crisp core region, and are in [0, 1] for those in the fuzzy boundary region. In other words, rough-fuzzy clustering provides a balanced mixture between *restrictive partition* of hard clustering and *descriptive partition* of fuzzy clustering. Therefore, it is faster than fuzzy clustering and is capable of better uncertainty handling/ performance (*IEEE Trans. SMC-B,* 37, pp. 1529-1540, 2007). Thus, wherever fuzzy c-means or c-medoids have been found to be useful in the past four decades, rough-fuzzy clustering would be superior. Merits of rough fuzzy integration have also been observed recently for designing a granular classifier with least dispersion index (*Pattern Recognition*, 45, pp. 2690-2707, 2012), case based reasoning classifier (*IEEE Trans. KDE,* 18, pp. 415-429, 2006) and in granular neural computation for feature selection (*Neural Networks,* 48, pp. 91-108, 2013).

Other salient contributions made are in - data condensation (*IEEE Trans. PAMI,* 24, pp. 1-14, 2002), unsupervised feature selection (*IEEE Trans. PAMI, 24,* pp. 301-312, 2002) and active support vector learning (*IEEE Trans. PAMI, 26,* pp. 413-418, 2004) in mining problems; stemming (*IEEE Trans. SMC-B,* 37, pp. 350-360, 2007) and page ranking in web intelligence

(*IEEE Trans. KDE,* 21, pp. 21-34, 2009); and gene selection (*IEEE Trans. SMC-B*, 40, pp. 741-752, 2010) and gene function prediction (*IEEE Trans. BE,* 56, pp. 229-236, 2009; *IEEE Trans. BE*, 59, pp. 1162-1168, 2012) in bioinformatics.

Computing With Words and Z-numbers: Current work

Let me now mention one of our current research problems, namely, computing with words (CWW) and the significance of the Z numbers, as recently explained by Zadeh (*Inform. Sci.*, 181, pp. 2923-2932, 2011). The CWW paradigm is inspired by the astounding ability of the human brain to perform tasks on the basis of concepts encoded in the words and phrases that frame natural language statements. It aspires to induce this amazing decision-making ability in the computer – a step towards evoking M-IQ (Machine-IQ) in a computer. CWW is imperative when a) the information to be conveyed lacks numeric precision; b) the situational imprecision can be exploited to arrive at robust, low-cost solutions; c) numeric computing principles cannot be applied; and d) words express a lot more than numbers. Potential areas of application of CWW are in semantic-web searching, linguistic summarization of text samples or complex phenomenon, and subjective decision-making.

The concept of the Z-number, proposed by Zadeh in 2011, depicts it as to being able to provide the perfect environment of amalgamation of CWW and behavioral computing. We have started investigating this issue (*Fundamenta Informaticae,* 124, pp. 197-229, 2013; *Theoretical Computer Science,* 448, pp. 2-14, 2013).

Summary

The aforesaid research in pattern recognition and machine intelligence *has led to the emergence of several modern disciplines* involving fuzzy sets with other computational paradigms, as evident from the literature. This has also led to the *incorporation of rough sets as a component of soft computing*; the original definition of Zadeh had components like fuzzy logic, neurocomputing, GA and probabilistic reasoning. This augmentation by rough sets has enhanced the computational intelligence property of soft computing and triggered its multifarious applications.

Graduated seventeen PhDs, till now, in India and abroad who are well placed in premier institutes and industry. Out of the ten PhDs from ISI, one is an SS Bhatnagar awardee, another is FNA, two are IEEE Fellow, four are FNAE and three are FASc, so far.

For furtherance of research in the said topics in a consolidated manner, our Institute created a new unit in March 1993, called Machine Intelligence Unit (MIU), under my headship. As recognition of our research contributions, the Department of Science and Technology, Govt. of India has established the nation’s first research center in Soft Computing in October 2004 at ISI, Calcutta under my leadership. The international conference PReMI (Pattern Recognition and Machine Intelligence) is the brain child of MIU. The conference is unique in the sense that it provides a platform to exchange ideas regularly between these two communities for mutual benefit. The first edition was in Calcutta in 2005, the fourth one in Moscow in 2011, and the fifth edition will be in ISI, Kolkata in 2013.

"Pattern Recognition" has not only led me win several prestigious awards and honors in India and abroad (including *S.S. Bhatnagar Prize* and *Padma Shri*), and attain the highest administrative and academic position, namely Director, of my Institute, but also gave me opportunities in visiting about forty countries for academic purpose and more importantly, to know the common people and culture there and make new friends. It will remain incomplete if I do not acknowledge my PhD students and young colleagues/ collaborators with whom I have done most of the work. They keep me always young.

References

[1] S.K. Pal and D. Dutta Majumder, *Fuzzy Mathematical Approach to Pattern Recognition*, John Wiley & Sons (Halsted), N. Y., 1986

[2] S.K. Pal and S. Mitra, *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing*, John Wiley, N.Y., 1999

[3] S.K. Pal and S.C. K. Shiu, *Foundations of Soft Case-Based Reasoning*, John Wiley, N.Y., 2004

[4] S.K. Pal and P. Mitra, *Pattern Recognition Algorithms for Data Mining*, CRC Press, Boca Raton, Florida, 2004

[5] S. Bandyopadhyay and S.K. Pal, *Classification and Learning Using Genetic Algorithms: Applications in Bioinformatics and Web Intelligence*, Springer, Heidelberg, 2007

[6] P. Maji and S.K. Pal, *Rough-Fuzzy Pattern Recognition: Application in Bioinformatics and Medical Imaging*, Wiley - IEEE, N.Y., 2012