Granular Mining and Big Data Analytics: Rough Models and Challenges

Sankar K. Pal

Proceedings of the National Academy of Sciences, India Section A: Physical Sciences

ISSN 0369-8203
Volume 90
Number 2

Proc. Natl. Acad. Sci., India, Sect. A Phys. Sci. (2020) 90:193-208
DOI 10.1007/s40010-018-0578-3

Springer

Proc. Natl. Acad. Sci., India, Sect. A Phys. Sci. (April–June 2020) 90(2):193–208
https://doi.org/10.1007/s40010-018-0578-3

REVIEW ARTICLE

Granular Mining and Big Data Analytics: Rough Models and Challenges

Sankar K. Pal[1]

Received: 1 June 2018 / Accepted: 18 December 2018 / Published online: 11 January 2019

Abstract Data analytics in granular computing framework is considered for several mining applications, such as in video analysis, bioinformatics and online social networks which have all the characteristics of Big data. The role of granulation, lower approximation and *r–f* information measure is exhibited. While the lower approximation over a video sequence signifies the object model for unsupervised tracking, it characterizes the probability (relative frequency) of definite regions in ranking miRNAs for normal and cancer classification. For neural learning, the information on definite region is used as the initial knowledge for encoding while generating the networks through evolution. Granules considered are of different sizes and dimensions with fuzzy and crisp boundaries. The tracking method is effective in handling different ambiguous situations, e.g., overlapping objects, newly appeared object(s), multiple objects in different directions and speeds, in unsupervised mode. The ranking algorithm could find only 1% miRNAs to result in significantly higher F-score than the entire set. Fuzzy–rough communities detected over the granular model of social networks are suitable in dealing with overlapping virtual communities in Big data. The knowledge encoding based on fuzzy–rough set provides superior performance than that of rough set. Future directions of research and challenges including the significance of *z*-numbers in precisiation of granules are stated. The article includes some of the results published elsewhere.

✉ Sankar K. Pal
sankar@isical.ac.in; sankarpal@yahoo.com

[1] Center for Soft Computing Research, Indian Statistical Institute, Kolkata 700108, India

Keywords Granular computing · Fuzzy–rough sets · Data mining · Bioinformatics · Video tracking · Social network analysis · Neural networks · *z*-Numbers · Granulated deep learning

1 Introduction

With the evolution of various modern technologies, huge amount of data are being constantly generated and collected around us. We are in the midst of what is popularly called information revolution and are living in a so-called world of knowledge. This huge amount of data, broadly characterized by three Vs—large volume, velocity and variety—is popularly known as "Big data." Analysis, access and storing of these data are now central to various scientific innovation, public health and welfare, public security and so on. Since Big data is highly complex in nature, mining them is not straight forward. Most of the information is heterogeneous, time varying, redundant, uncertain and imprecise. To reason, understand and mine the useful knowledge from these data have become a great challenge.

Big data analytics (BDA), i.e., analytics over Big data, has tremendous promising role in several fields such as commerce and business, biology, medicine, public administration, material science and cognition in human brain, just to name a few. Its objective, as it stands, is to develop complex procedure running over large-scale enormous-sized data repositories for—extracting useful knowledge hidden therein and delivering accurate predictions of various kinds within a reasonable time period. While analytics over big data has a leading role, there has been a significant shortage of deep analytical talent globally.

Mining of Big data can be made effective with methodologies that can deal with their characteristics, such as heterogeneity, dynamism or time varying, redundancy, uncertainty and impreciseness. Heterogeneity in data comes from the characterization of information in different ways, such as using numerical, categorical, text and image or audio/video data. Dynamism in data is due to the mechanism that generates related data changes at different times or different circumstances, which adds new uncertainty and difficulties for analysis. High dimensionality is due to the inappropriate collection of data for a task. Although a large candidate set of attributes is provided, most of them are often irrelevant or redundant. These superfluous attributes deteriorate the learning performance of decision-making algorithms.

Granular computing (GrC) is a nature-inspired information processing paradigm [1] that works with the process of *information granulation/abstraction,* and where the computation is performed using *information granules* rather than the data points (objects). In the recent past, GrC has evolved as a forefront research area due to the need and challenges from various domains of applications, such as data mining, document analysis, financial gaming, organization and retrieval of huge data bases of multimedia, medical data, remote sensing and biometrics. GrC, based on technologies like fuzzy sets [2], rough sets [3, 4] and computing with words [5], provides powerful tools for multiple granularity and multiple-view data analysis, which is of vital importance for understanding the complexity in Big data. There are some other applications which may require understanding of Big data at different levels of granularity. That means, GrC has significance both in processing and decision-making levels.

The present article describes the significance of granular computing in several mining applications, such as in video analysis, bioinformatics and online social networks which are considered to have all the aforesaid characteristics of Big data. Section II briefly mentions the salient features of GrC and the relevance of fuzzy sets and rough sets therein. Section III explains the concept of lower and upper approximations of a cluster (class of patterns) in the framework of rough set theory and the associated uncertainty. Sections IV and V describe the generalized rough set and rough information granule, respectively. Generalized rough set involves the incorporation of the concept of fuzziness into its set and/or granules and provides a stronger paradigm for handling uncertainties arising from overlapping classes/concepts as well as from granularity in the domain of discourse. On the other hand, rough information granules have the inherent characteristics of dimensionality reduction. Sections V to VIII demonstrate the application of GrC in problems like unsupervised video tracking, miRNA ranking for cancer detection and community detection in social networks, respectively. Here the roles of granule, rough lower approximation and rough–fuzzy information measure are highlighted. The unsupervised tracking system uses various neighborhood granules as obtained from the object model which, in turn, is extracted from the lower approximation of video frames. The system has the capability of handling efficiently different ambiguous situations, like tracking multiple objects moving in different directions with different speeds, overlapping objects, objects of similar color and newly appeared object while tracking some others. In the problem of miRNA ranking, fuzzy lower approximations of normal and cancer classes are used to find the probability (relative frequency) of definite and doubtful regions for entropy computation. A granular model, named fuzzy granular social network (FGSN), is illustrated in Section VIII for community detection. Section IX explains, in short, the basic principles of designing granular neural networks based on information granules, as extracted from lower approximation of classes. The merits of fuzzy–rough sets over rough sets in encoding knowledge are demonstrated. Section X finds conclusions and future directions of research.

2 Granular Computing: Why Rough and Fuzzy Sets?

Granular computing (GrC) has basically three components, namely granulation, granules and computing with granules. Granulation means natural clustering which replaces a fine-grained universe by a coarse-grained one, more in line with human perception. This can be viewed as a process like—self-organization, self-production, morphogenesis and Darwinian evolution—that is abstracted from natural phenomena. Clusters or segments so formed by granulation (natural clustering) are called granules. That is, a granule may be defined as a clump of indiscernible entities (drawn together, say, by similarity, proximity or functionality) with respect to given attributes. Granulation is therefore a process of formation and representation of granules.

Granular computing (GrC), as stated before, provides an information processing paradigm that works with the process of *information granulation/abstraction,* and where computation is performed using *information granules* and not the data points (objects). Since granules refer to compressed information and operations are performed on them, GrC leads to have both data compression and gain in computation time, and therefore finds wide applications in data analytics.

In many situations, when a problem involves incomplete, uncertain and vague information, it may be difficult to differentiate distinct elements and one may find it convenient to consider granules [1]. On the other hand, in some applications though detailed information is available, it may be sufficient to use granules in order to have an efficient and practical solution. From a more practical point of view, the simplicity derived from granular computing is useful for designing scalable data mining algorithms.

Further, varying the size and shape of granules, one may determine different levels of granularity, characterize a specific aspect of the problem, represent the model differently and finally regulate the decision-making tasks by utilizing them for problem solving. Therefore, formation of granules plays a significant role in GrC. For further reference, one may cite [6, 7].

It may be noted that the concept of granulation is inherent in theories of both fuzzy sets (FS) and rough sets (RS). Membership function (which is the building block of fuzzy set theory) practically granulates the features, thereby producing fuzzy granulation of feature space. On the other hand, rough set theory concerns with a granulated domain where a crisp set is defined thereon. Subsequently, fuzzy sets and rough sets became most successful technologies for GrC. Between the two, rough set theory has enriched the literature of GrC significantly. This is evident also from a recent survey [8].

While the theory of fuzzy sets is pivoted on the concept of membership function characterizing the degree of belonging of an element to a set representing an imprecise concept, the rough set theoretic approach is based on the principles of granular approximation of a set from its inner and outer sides concerning the belonging of granules to it. These approximations which represent the groups of granules 'that definitely', and 'definitely and possibly' belong to the set are called lower and upper approximations, respectively. A nice tutorial on rough sets and its rudiments can be found in [9, 10].

Let us now explain the meaning of lower and upper approximations of a cluster or a pattern class and the associated uncertainty for machine learning problems.

3 Lower and Upper Approximations of a Cluster

Let us consider a cluster β defined over a granulated domain (Fig. 1). Here white granules inside the cluster, which have no doubt of their belonging to it, constitute what is called *lower approximation* of β. The shaded granules, on the other hand, are the boundary granules about which we have doubt on their belonging to the cluster β. These boundary granules together with those in the lower approximation constitute what is called *upper approximation* of β. Roughness of β is defined as:

$$\text{Roughness of } \beta = 1 - |\text{Lower}|/|\text{Upper}| \quad (1)$$

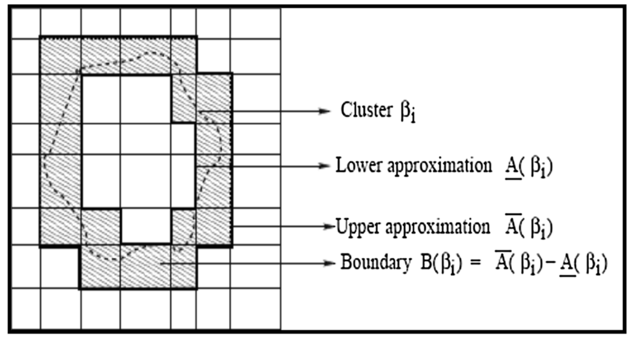

Fig. 1 A cluster β with lower and upper approximate regions

|Lower| and |Upper| denote the number of granules in lower and upper approximations, respectively. When |Lower| = |Upper|, $\beta = 0$. That is, if there is no granularity in the domain, there is no roughness in the cluster.

In Fig. 1, both the cluster and the granules are crisp. However, in real life either or both of them could be fuzzy. In that case, the lower and upper approximate regions would also be fuzzy, characterized by respective membership functions. This leads to the definition of *Generalized Rough Sets* which provides a stronger model of uncertainty handling. It is stronger in the sense that it models the uncertainty arising from both overlapping character of clusters and granularity in the domain.

Let us now explain the lower and upper approximations of generalized rough sets [11].

4 Generalized Rough Sets: Lower and Upper Approximations

In Pawlak's rough set theory [3], both the set X and granules or equivalence relation R are considered to be crisp. However, in real-life problems, they could be fuzzy too. Generalized rough sets are defined based on this premise where the expressions for the lower and upper approximations of a set X depend on the type of relation R and whether X is a crisp or a fuzzy set. These are described as follows:

Case 1 R denotes a crisp equivalence relation and X is a crisp set. The expressions for the lower and upper approximations of the set X are given as

$$\begin{aligned} \underline{R}X &= \{u | u \in U : [u]_R \subseteq X\}, \\ \overline{R}X &= \{u | u \in U : [u]_R \cap X \neq \emptyset\}, \end{aligned} \quad (2)$$

Here $[u]_R$ denotes the granule to which the element u belongs. The pair of crisp sets $\langle \underline{R}X, \overline{R}X \rangle$ is referred to as the rough set of X and $\langle U, R \rangle$ is a crisp equivalence approximation space.

Case 2 R denotes an equivalence relation and X is a fuzzy set. The lower and upper approximations of the set X are expressed as

$$\underline{R}X = \left\{ \left(u, \inf_{z \in [u]_R} \mu_X(z) \right) | u \in U \right\},$$
$$\overline{R}X = \left\{ \left(u, \sup_{z \in [u]_R} \mu_X(z) \right) | u \in U \right\}, \tag{3}$$

Here μ_X is the membership function associated with X. The pair of fuzzy sets $\langle \underline{R}X, \overline{R}X \rangle$ is referred to as the rough–fuzzy set of X and $\langle U, R \rangle$ is a crisp equivalence approximation space.

Case 3 R refers to a fuzzy equivalence relation, that is, the belongingness of every element (u) in the universe (U) to a granule $Y \in U/R$ is specified by a membership function, say, m_Y, that takes values in the interval $[0, 1]$ such that $\sum_Y m_Y(u) = 1$. In such a case, when X is a crisp set, the lower and upper approximations of X are defined as

$$\underline{R}X = \left\{ \left(u, \sum_{Y \in U/R} m_Y(u) \times \inf_{\varphi \in U} \max(1 - m_Y(\varphi), C) \right) | u \in U \right\},$$
$$\overline{R}X = \left\{ \left(u, \sum_{Y \in U/R} m_Y(u) \times \sup_{\varphi \in U} \min(m_Y(\varphi), C) \right) | u \in U \right\}, \tag{4}$$

where

$$C = \begin{cases} 1, & \varphi \in X \\ 0, & \varphi \notin X \end{cases} \tag{5}$$

Symbols $\sum$ (sum) and $\times$ (product), respectively, represent specific fuzzy union and intersection operations. (One may consider any fuzzy union and intersection operation instead of the sum and product operations depending on applications.) The pair of fuzzy sets $\langle \underline{R}X, \overline{R}X \rangle$ is referred to as the fuzzy rough set of X and $\langle U, R \rangle$ is a fuzzy equivalence approximation space.

Case 4 In *Case 3* of R referring to a fuzzy equivalence relation, when X is a fuzzy set, the expressions for the lower and upper approximations of X are

$$\underline{R}X = \left\{ \left(u, \sum_{Y \in U/R} m_Y(u) \times \inf_{\varphi \in U} \max(1 - m_Y(\varphi), \mu_X(\varphi)) \right) | u \in U \right\}$$
$$\overline{R}X = \left\{ \left(u, \sum_{Y \in U/R} m_Y(u) \times \sup_{\varphi \in U} \min(m_Y(\varphi), \mu_X(\varphi)) | u \in U \right) \right\}. \tag{6}$$

The pair of fuzzy sets $\langle \underline{R}X, \overline{R}X \rangle$ is referred as the fuzzy rough–fuzzy set of X and $\langle U, R \rangle$ is again a fuzzy equivalence approximation space.

It is clear from the above discussion that the expressions for lower and upper approximations in Cases 1–3 are special cases of those of Case 4. Pictorial diagram of lower and upper approximations for Case 4 is shown in Fig. 2.

Given the fuzzy lower and upper approximations of the generalized rough sets X, one can determine the roughness value of X with Eq. (1) where |Lower| and |Upper| would denote fuzzy cardinalities. Using logarithmic gain function or exponential gain function, an entropy measure can be defined [11] based on roughness measures of X and its complement X^c in order to quantify the incompleteness of knowledge about a universe.

This entropy measure can be used for mining and analyzing any kind of data, be it, say, biological, or astronomical or video, wherever the uncertainty arises from overlapping concepts/classes, as well as from granularity in the domain of discourse. The uncertainty due to randomness in the occurrence of events has not been considered here. Significance of generalized rough sets and entropy in defining various image ambiguity measures for image analysis is described in [11, 12].

In the following section, we explain *f*-information granulation using rough sets and its significance in dimensionality reduction. Then, we demonstrate some applications of granular computing in handling three types of Big data problems, namely miRNA selection in bioinformatics, video tracking and social network mining. These are followed by granular neural net design for efficient learning.

5 Rough F-Information Granulation and Dimensionality Reduction

Development of information granules using rough set theory has been made by several researchers. The theory is used in [9] to obtain dependency rules which model various informative regions in fuzzy granulated feature space. The fuzzy membership functions corresponding to these informative regions constitute what are called rough

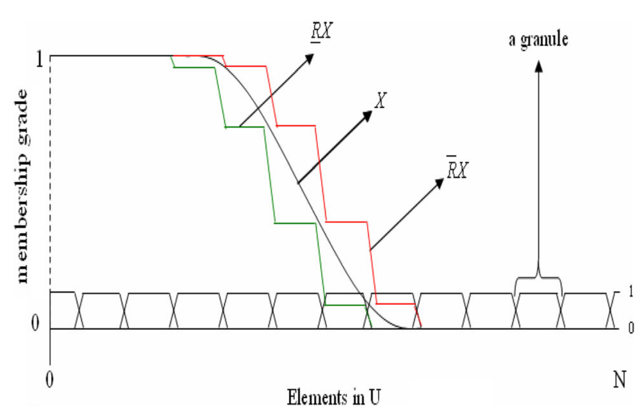

Fig. 2 Pair $\langle \underline{R}X, \overline{R}X \rangle$ is referred to as fuzzy rough–fuzzy set of X

information granules. These granules may involve a reduced number of relevant features, thereby resulting in dimensionality reduction. For example, given the object region in Fig. 3, rough set theory can, whether supervised or unsupervised mode, extract the rule $F_{1M} \vee F_{2M}$ (i.e., feature F_1 is M AND feature F_2 is M) to encode the object region. This rule which represents the rectangle (shown by bold line) provides a crude description of the object region and can be viewed as its information granule. Though both the features F_1 and F_2 have appeared here in the rule, in practical scenario with large number of features, the rough rules corresponding to various regions may not contain all of them.

For example, consider IRIS data with three flowers viz, Setosa, Versicolor and Virginica having, four features viz, sepal length, sepal width, petal length and petal width, and 50 samples from each category. From the scatter plots taking two features together, two flower classes are seen to be overlapped while the other one well separated. So, intuitively one may need all the four features to segregate mostly the ovelapping flowers, but definitely not for the other flower which is well separated. That is, on an average the number of features required to characterize/represent a flower class would be less than four. This is what is reflected in Fig. 4 where the aforesaid rough–fuzzy information granulation needs only about 2.5 features on average to store a flower (case) from each class instead of 4, as required by IB3, IB$ and random selection methods [13]. Even with this reduced number of features, the classification accuracy by 1-NN classifier is maximum. The retrieval time (tret), as expected, is minimum. Generation time (tgen) is next to random selection method which needs almost no time to pick up a representative flower from each category.

Its application in information retrieval for mining large data sets has been adequately demonstrated [13]. In [14], the concepts of rough sets are also employed for generating multi-granulations, based on multi-equivalence relations. A neighborhood type of information granule is developed in [15]. Here, a neighborhood rough set defines the granule and its size is controlled by a distance function. Information granulation using covering-based rough set is provided in [16].

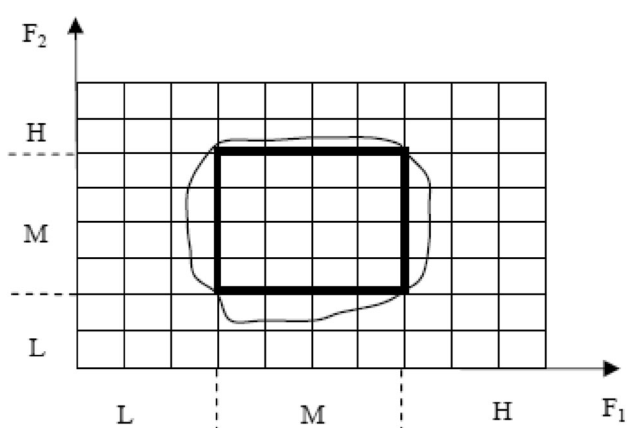

Fig. 3 Rough set theoretic rules for an object. L, M and H represent low, medium and high, respectively

In the following sections, we demonstrate the development of analytics using GrC for three types of Big data, namely videos, miRNA expressions and online social network data. Example tasks considered are unsupervised video tracking, miRNA ranking in cancer detection and community detection, respectively. The challenging issues in the respective domains are addressed. Apart from performance, the results exhibit the role of

- Granules
- Lower approximation
- *r–f* information (entropy) measure.

6 Video tracking

Tracking of moving object(s) from video sequences is one of the basic steps of video processing [17–19]. Object tracking is required in several fields of computer vision, e.g., surveillance and gesture recognition [18]. This problem has been studied over the decades. Some of the approaches are partially supervised where the initial object/background model is known, while some others are unsupervised. One may note that in video tracking the complete information is not available. This makes the prediction task difficult as it involves several uncertainties due to, e.g., change in shapes/sizes of moving object(s), change in motion of the object(s) and change in numbers of object(s).

Here we describe an unsupervised method of tracking [19] using adaptive rule generation in granular computing framework based on rough rule base and granular flow graph. The flow graph enables in defining an intelligent technique for rule base adaptation where its characteristics in mapping the relevance of attributes and rules in decision-making system are exploited. These techniques are performed in neighborhood granular level. The rough flow graph-based adaptive granular rule-based system, produced for unsupervised video tracking, is able to overcome the incompleteness in information that arises without the initial manual interactions and in providing superior performance while gaining in computation time. The cases of partial overlapping of objects and their unpredictable changes are handled efficiently.

The major characteristics of the method are as follows:

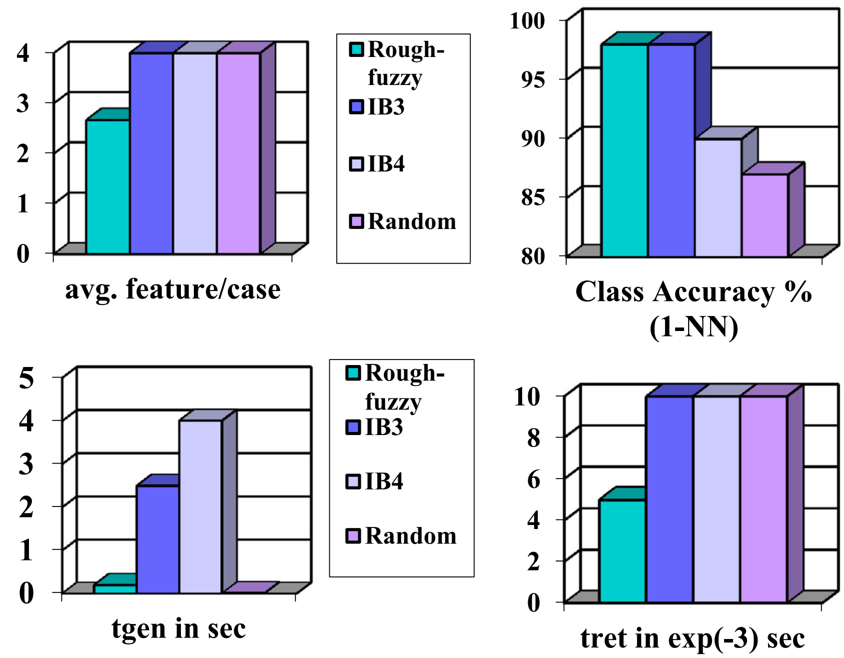

Fig. 4 Performance of different case generation schemes for the IRIS data with 3 classes, 4 features and 150 samples

- Neighborhood granules in RGB-D space are used. Granules are of three types, viz, 3-d spatiotemporal, 2-d spatio-color, 1-d color, which are formed by region growing using color and spatial similarity of pixels. Granules thus generated are of arbitrary shape, i.e., natural granules.
- Granules are extracted from *Lower Approximate Object Model* which is obtained in the *median frame* using temporal information in D-space.
- Involves granular level rule-based decision, i.e., uses granules, instead of pixels, for object-background (O/B) partition.
- Updating of rule base is done automatically with flow graph.
- Deals with ambiguous tracking situations (e.g., overlapping objects, newly appeared object, objects merged with similar color, multiple objects moving in different directions and speed) in unsupervised mode.

6.1 Lower Approximate Object Model

Let us now explain the lower approximate object model.

Given P frames $f_t, f_{t-1},\ldots f_p,\ldots f_{t\ -P-1}$ of a video, compute the absolute difference frames $\{\tau_p\ \forall\ p \in 1,2,\ldots,P-1\}$ w.r.t. f_t in D-space to find temporal (changed) information. Then, performing intersection ($\cap$) and union ($\cup$) operations as follows

- $\cap\{\tau_p\ \forall\ p \in 1,2,\ldots,P-1\}$ denotes the common moving regions of τ_p over $P-1$ frames and provides an estimation of *Lower approximation* (O_{low}) of the moving object(s) in temporal domain. This may be termed as the object model in D-space.
- $\cup\{O_{\mathrm{low}},\ \tau_P\} = \cup\{O_{\mathrm{low}},\ |\ f_t - f_{P-1}|\}$ (basically τ_P) denotes an estimation of *Upper approximation* (O_{up}) of the moving object(s) in temporal domain.

Values of the features (RGB-D, Temporal) contained in the set O_{low} are the core values of the object model. Values of these features in the set $\{O_{\mathrm{up}} - O_{\mathrm{low}}\}$, i.e., the boundary region, determine the extent to which the values in the object model are allowed. These approximations are shown in Fig. 5 for $P = 4$ where a girl is doing exercise by moving her hands.

6.2 Formation of Granular Rule Base

Formations of 2-d spatio-color granules, 1-d color granules and finally 3-D spatiotemporal granules are shown in Fig. 6. First, 2-D spatio-color granules are extracted in the lower approximate object model in τ_{median} frame. Then, 1-d color granule corresponding to the representative location of a 2D spatio-color granule is formed using the temporal information from the $\tau_{\mathrm{P}}\ldots\tau_1$ planes, where $\tau_P = |f_t - f_{P-1}|$. These two types of granules together constitute the 3-d spatiotemporal granules of the object model. These are called neighborhood granules as they use neighborhood pixel information while forming.

In Fig. 6, three such 1-d granules consisting of 2, 3 and 1 points (as shown connected by solid lines) are marked as examples. Once these points are obtained in τ-plane, find the corresponding point locations in f-plane and their RGB

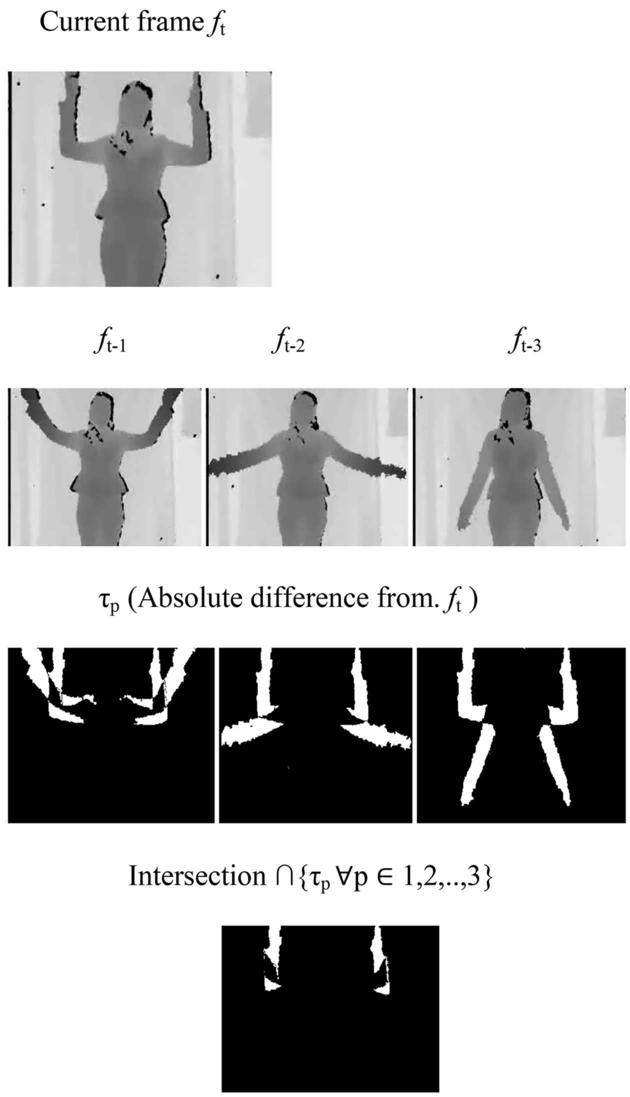

Fig. 5 Extraction of temporal information and lower approximate object model (in D-space, $P = 4$)

and D values. Given the {RGB, D, τ} feature information of a1-d granule G_c, determine the range of each of these features considering a possible deviation δ, say equal to 2. These three ranges, say RGB_N, D_N and $Sp\text{-}Temp_N$, are then presented to the rule base as input to characterize the range of the object model with respect to G_c.

6.3 Methodology

The block diagram of the method of unsupervised tracking is shown in Fig. 7. Here the rule base is updated using a granular flow graph. It is called granular flow graph as it maps the decision paths of the granular rule base. The method of updating is intelligent in the sense that neither the ranges of all the features for an object, nor the ranges of all objects need to get updated at a go (simultaneously).

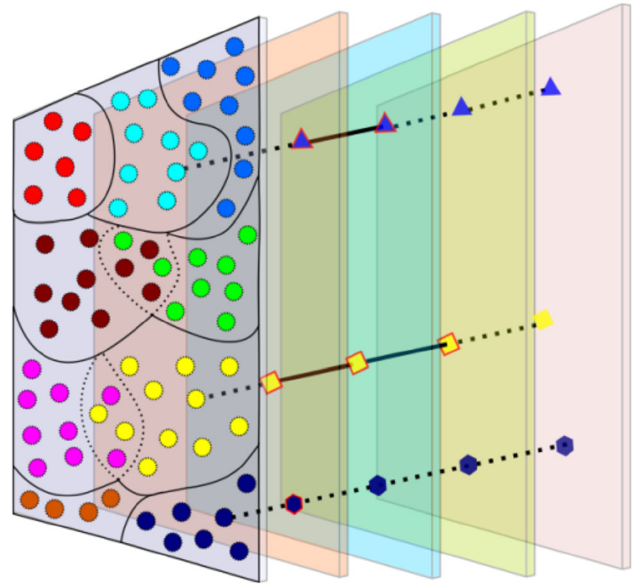

Fig. 6 3-d spatiotemporal granules extracted in lower approximate object model in τ_{median} frame

Updating is done for a particular feature(s) *only* when required, instead of every time. For details on the functioning of flow graph, one may refer [20].

Figure 8 demonstrates, as an example, the capability of the system in tracking multiple objects moving in different directions with different speeds. Here the frames per sec = 15, and $P = 6$. Similarly the system could handle efficiently the ambiguous situations like tracking overlapping objects, objects of similar color and newly appeared object while tracking some others. Further, as it uses granules rather than pixels for object-background partition, the tracking is fast. The details are available in [19]. Other approaches of video tracking in granular computing framework, e.g., using fixed rule base without flow graph, using rough entropy with spatiotemporal granules and using neighborhood rough filter and intuitionistic entropy, are reported in [21–23]. Readers may also consult [24–34] for some important tracking algorithms based on technologies, other than granular computing.

7 Micro-RNA Ranking in Cancer Detection

Let us now consider an important problem in bioinformatics, namely identifying MicroRNAs (miRNAs), which are responsible for cancer detection. MicroRNAs are non-coding RNAs, and they work on messenger RNAs (mRNA) to inhibit protein translation by degrading the mRNAs. They act as a major biomarker of cancer. However, all miRNAs in human body are not equally important for cancer identification, and the role of various miRNAs is

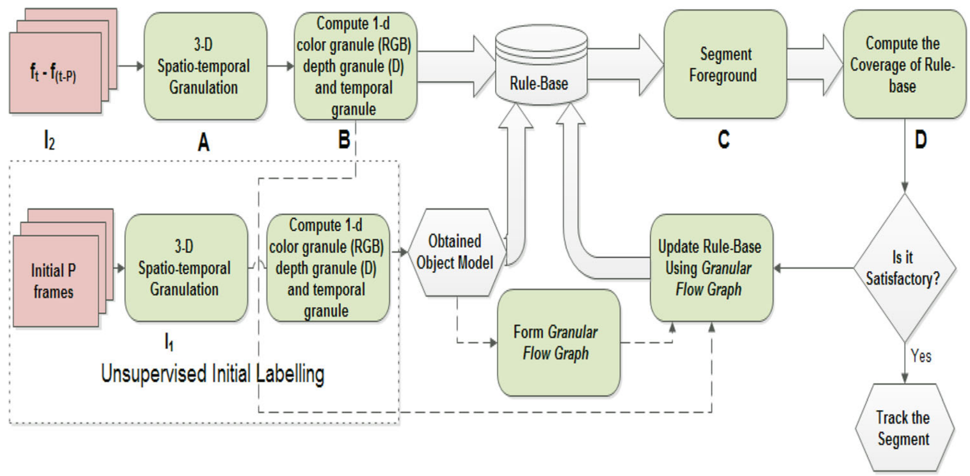

Fig. 7 Block diagram of unsupervised tracking

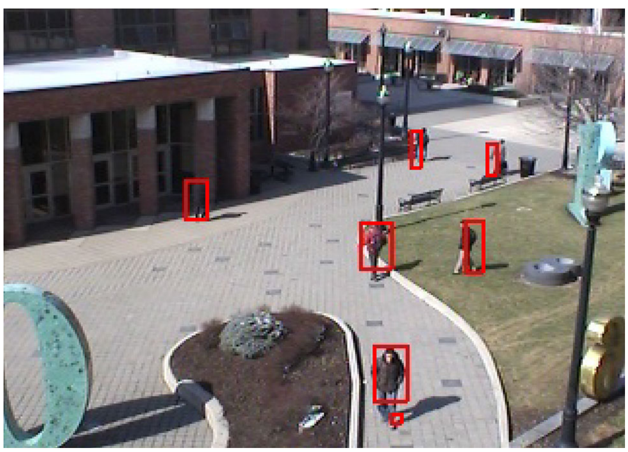

Fig. 8 Tracking multiple persons moving in different directions and speeds

diverse for different types of cancers. Further, the presence of any irrelevant miRNA may decrease the classification accuracy and increase both the biochemical and computational costs. Therefore, selection of most informative miRNAs is important for identifying the condition of a sample/patient.

It may further be mentioned that in miRNAs expression data, the number of training samples is usually very small compared to large number of miRNAs involved in the experiments. And among this large amount, only a small fraction is effective for a certain task. So, the process of selecting informative miRNAs concerns basically a small sample, large dimension problem, unlike the conventional feature selection techniques in pattern recognition and machine learning where the number of samples is very large than that of features.

Here we describe the salient features and results of a recent method [35] for identifying a reduced set of most relevant miRNAs through their ranking. The method uses a fuzzy–rough entropy measure (FREM) to determine the relevance of a miRNA in terms of separability between normal and cancer classes. While computing the FREM for a miRNA, fuzziness takes care of the overlapping between normal and cancer expressions, whereas rough lower approximation determines their class sizes. MiRNAs are sorted according to the highest relevance (i.e., the capability of class separation) and a percentage among them is selected from the top ranked ones. FREM can also be used to determine the redundancy between two miRNAs for removal of the redundant ones from the aforesaid selected set, if required.

Salient features of the algorithm are as follows:

In the framework of generalized rough sets (Sect. 4), here the set is considered to be crisp (i.e., either cancer *C* or normal *N*) and the granules fuzzy. Therefore, the associated entropy is called *fuzzy–rough entropy*.

Fuzzy Lower Approx. of *N* and *C* classes are used to find the probability (relative frequency) of definite and doubtful regions for entropy computation.

Entropy minimization implies higher relevance of a miRNA.

Top 1% selected miRNAs provide significant improvement over the entire set in terms of F-score.

Before the results are described, we explain the FREM with respect to a miRNA.

Fig. 9 1-d fuzzy granules of normal (N) and cancer (C) classes for a miRNA

7.1 Fuzzy–Rough Entropy

Figure 9 shows, as an example, the pictorial diagram of 1-d fuzzy granulation formed by normal and cancer patients with respect to a miRNA, say M_1 where the decision region is crisp, i.e., either normal or cancer.

Let $\mu_N(p)$ be the membership values of a patient p to normal class (N) with respect to a miRNA M_1, and $\mu_C(p)$ be the same for cancer class (C) where $\mu_N(p) + \mu_C(p) = 1$. Then, the membership values of p to lower approximate regions of N and C are defined as in [35].

$$\mu_{\mathrm{low}N}(p) = \min\{\mu_N(p),\ \theta\}, \theta = 1 \quad \text{if} \quad p \in N, \quad \text{and } 0 \quad \text{if} \quad p \in C$$
$$\mu_{\mathrm{low}C}(p) = \min\{\mu_C(p),\ \theta\}, \theta = 1 \quad \text{if} \quad p \in C, \quad \text{and } 0 \quad \text{if} \quad p \in N$$

$\mu_{\mathrm{lowN}}(p)$ and $\mu_{\mathrm{lowC}}(p)$ denote the degree of being sure to be a N patient and C patient, respectively.

Relative frequencies (normalized cardinalities) of these lower approximate regions are therefore:

$$\lambda_{\mathrm{low}N} = (1/|N|)\Sigma_p \mu_{\mathrm{low}N}(p)$$
$$\lambda_{\mathrm{low}C} = (1/|C|)\Sigma_p \mu_{\mathrm{low}C}(p)$$

Here $|N|$ and $|C|$ denote the total number normal and cancer patients, respectively.

One may then compute the following:

Mean relative frequency of the lower approximate regions of N and C as

$$\lambda_1 = (\lambda_{\mathrm{low}N} + \lambda_{\mathrm{low}C})/2$$

Relative frequency of the overlap region between N and C as

$$\lambda_2 = 1 - \lambda_1$$

Fuzzy–rough entropy of the system with respect to miRNA-M_1

$$\mathrm{FREM}_1(N,\ C) = -\sum_i \lambda_i \log 2\lambda_i,\ i = 1,\ 2.$$

7.2 Relevance of a miRNA

Decrease in FREM_1 implies increase in the distance between low_N and low_N regions, i.e., increase in separation of N and C classes, and hence the increase in relevance of the miRNA M_1 for cancer classification.

Increase in the overlap of N and C regions implies increase in the closeness of λ_1 and λ_2 values, i.e., increase in FREM_1.

FREM_1 attains the maximum value (= 1) when $\lambda_1 = \lambda_2 = 0.5$.

7.3 Results

The experiment involves computation of relevance of each miRNA using the aforesaid FREM_1 measure, ranking them based on this, and selecting a set of most relevant ones, as per the necessity. Figure 10 provides some sample results in terms of F-score obtained by 1% miRNAs, thus selected, for six normal-cancer data sets when the classification was done using SVM (support vector machine). Results are compared with those obtained by the entire sets of miRNAs. Higher value of F-score implies higher possibility of detecting cancer patients as cancer and normal patients as normal, as it should be.

As seen from the figure, only 1% selected genes result in significantly higher F-score than the entire set. This means, that 1% is only effective while the others are irrelevant as far as cancer detection is concerned. These irrelevant miRNAs not only affect the classification accuracy, but also increase the computational and other costs.

For example, the breast, colorectal, lung, melanoma, pancreas and nasopharyngeal data sets have total number of 309, 352, 866, 864, 847 and 887 miRNAs, respectively. Out of them, only 10, 7, 4, 6, 16 and 5 number of selected ones are effective for detecting the respective cancers.

Size of the set of miRNAs, thus selected based on their relevance, can be further reduced, if required, by removing the redundancy, if any, within its constituting miRNAs. Redundancy of M_i with respect to another miRNA M_j, $i \neq j$, can be computed similarly by the aforesaid FREM measure where one needs to consider M_1 and M_2 as two separate classes instead of N and C, and the number of samples in each of those classes would be equal to the total number of N and C patients. Here lower value of FREM_i $(M_i,\ M_j)$ $i \neq j$ indicates larger separability (i.e., larger redundancy) of M_j with respect to M_i, and hence stronger

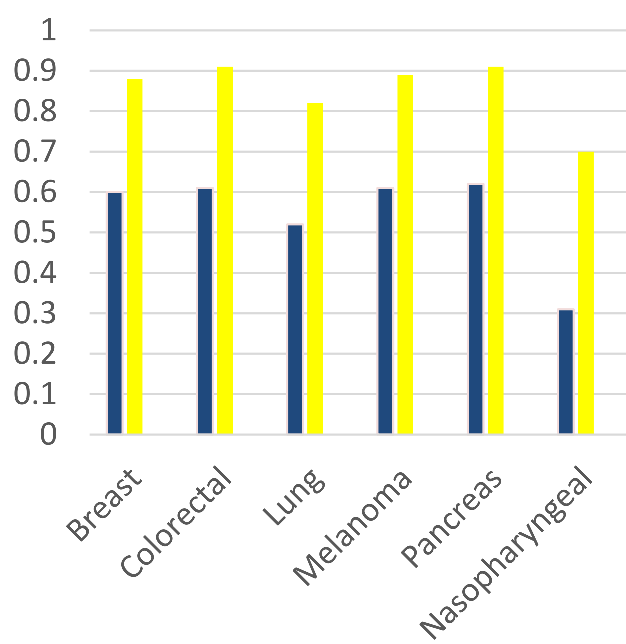

Fig. 10 F-scores for six types of normal-cancer data sets

likelihood for removal of M_i. The details of the methodology can be found in [35].

The aforesaid algorithm of selecting the relevant miRNAs is based on the ranking of miRNAs as per their individual relevance. Methods are also available [36, 37] where the selection is made depending on the relevance of miRNAs in a group, rather than their individual ranking.

Since the literature on gene selection is much richer than that of miRNA, readers may refer to [38–46] which include some relevant methods and measures for gene selection as well as some well-known algorithms for feature selection in large data sets, in general.

8 Community Detection in Social Networks

Let us now consider another example of Big data, namely online social networks which is highly voluminous, dynamic and complex. Detecting communities in such a data is a challenging task. This problem is addressed here in granular computing framework.

Social networks have following characteristics [47]:

- The degree distribution of social networks is skewed, following the power law or truncated geometric distribution.
- Diameter of the social network is found to be very small compared to the size of the network, and
- Social network possesses high concentration of edges in its certain parts forming groups.

This last feature, i.e., forming groups with high internal edge density within themselves and low between them characterizes the community structure (or clustering) of the network.

8.1 Why Community Detection?

In a society, one can find groups that naturally form, such as families, co-workers' circles, friendship circles, villages and towns. Similarly, 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 helps in optimizing the internet infrastructure. In a purchase network, this can boost the sell by recommending appropriate products. In computer network, it helps to optimize the routing table creation. In citation network, it finds researchers of similar fields. Detecting these communities also helps in identifying the special actors. The central nodes of the clusters or nodes in the boundary region (which act as a bridge between communities) are called special actors as they play different important roles within the society.

8.2 Why Fuzzy Granules?

A social network is viewed as a collection of relations between social actors (nodes) and interactions between social actors. These actors are often indistinguishable in some problem-solving tasks. This signifies the relevance of dealing them with the concept of *Granules.* Further, the relations/interactions between nodes and clusters of nodes do not often lend themselves to precise definition. For example, in friendship networks, suppose x and y were initially direct friends. Later on another person z became a friend of x through y. In that case, the degrees of friendship between *x* & *y* and *x* & *z* may not be the same, though they are all friends. This can be modeled in the notion of fuzzy set theory by describing *Fuzzy* granules and boundaries.

Consider, for example, the Zachary Karate Club network (Fig. 11). This network shows the friendship relations between 34 members of a US Karate club in the 1970s (links indicate 'friendship' between club members). Let us now construct a granule with (virtual) radius of 1 *hop* distance around each node (or actor). The actor around which the granule is constructed is referred as the center (representative) of that granule. Because of the dense structure of the network, these granules would be overlapping.

If we consider such granules centered at nodes 1 and 3, one can see that nodes such as 4, 8, 2 and 14 belong to both the granules. Fuzzy membership can therefore be used to represent the degree of multi-class belonging of nodes.

Based on the aforesaid concept of granulation, undirected social networks can be represented by a triple:

$$S = (C, V.G)$$

where V is the finite set of nodes of the network, $C \subseteq$ V is a finite set of granule representatives and G is the finite of all granules.

This model is called fuzzy granular social network (FGSN) [48].

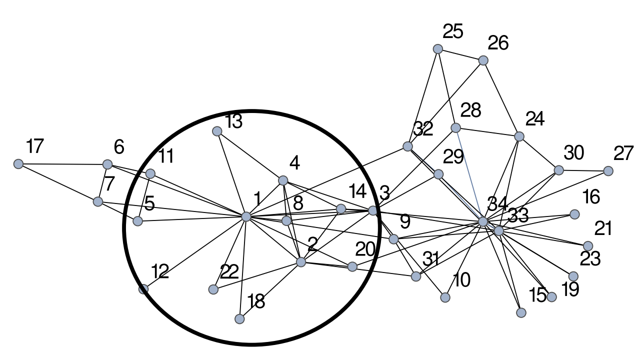

Fig. 11 Zachary karate club network with a virtual granule of radius of one *hop* distance around node no 1

8.3 Fuzzy–Rough Community Detection in FGSN: Principle

A community is formed when nodes are densely connected compared to the other parts of the network. In the knowledge representation of FGSN, one may find out densely connected granules. The network in Fig. 11 appears to have two communities, likely overlapping.

The basic steps of detecting communities in FGSN are [49]:

- Identify the dense granules.
- Merge them if they are nearby.
- Form a meaningful fuzzy community by discarding the weakly coupled granules.
- Form fuzzy–rough representation of the output community structure.

Dense granules are those whose granular degree exceeds a threshold. Merging enables identifying the dense granules belonging to the same community. Weakly coupled granules are those whose granular embeddedness is less than a threshold.

Communities, so identified, have fuzzy boundaries. Rough set theoretic representation of a fuzzy community structure means the following: The nodes belonging to lower approximations of the constituting communities have membership value $\mu = 1$, as they definitely belong to the respective communities. Nodes in boundary (i.e., upper–lower approximations) region have membership value $0 < \mu < 1$, as they possibly belong to the communities. The output thus generated is called *Fuzzy–rough Community*.

Note that, in the process one may generate orphans. A node is said to be an orphan if it not a member of any community. The details of the method are explained in [49].

8.4 Results

Figure 12 shows some example results demonstrating the effectiveness of the algorithm on LFR benchmark data of a undirected friendship network variation of the normalized fuzzy mutual information (NFMI) between two sets of communities with mixing parameters η and O_n, which reflect different degrees of overlapping between them, is exhibited. The figure shows a comparison of the aforesaid algorithm (FRC-FGSN) with three well-known graph theoretic methods.

NFMI measures [49] the goodness of a detected community structure, given the actual one. Higher value of NFMI signifies larger similarity (or relevance) between two communities. η denotes the average fraction of edges shared with other community with respect to those within communities. O_n represents the fraction of nodes belonging to overlapped regions. Therefore, overlapping between communities increases with the increase in their values.

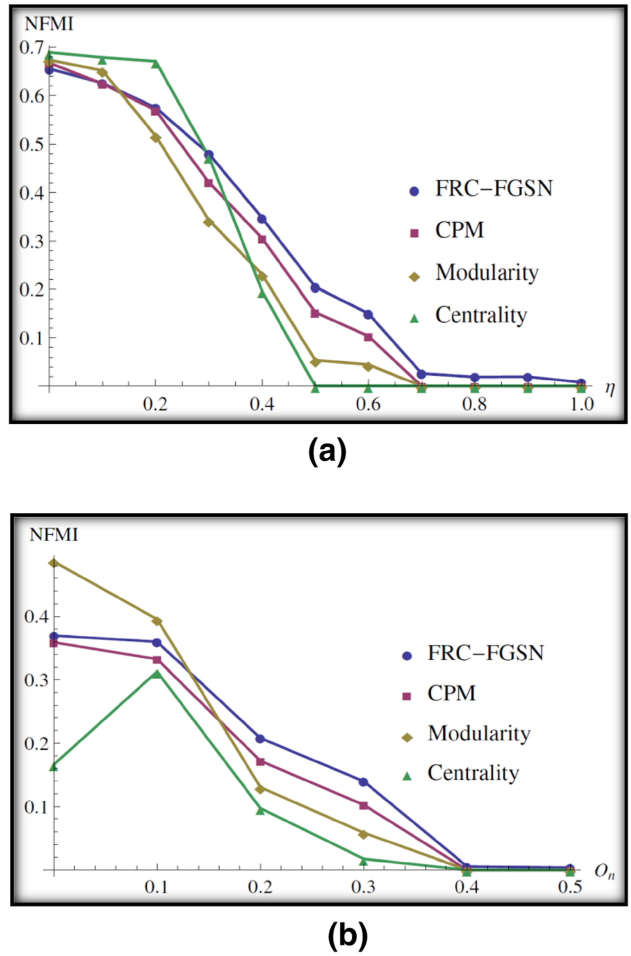

Fig. 12 Variation of NFMI with **a** η and **b** O_n

Centrality-based method [50] and modularity optimization method [51] for community detection produce crisp (disjoint) communities. *k*-clique percolation method [52] produces overlapping communities, but a node in overlapped region has "full" membership to each of the intersecting communities.

For a given set of η, O_n, network size, and maximum and minimum number of nodes in a cluster (i.e., community size), a network with specified (known) number of clusters is generated in LFR benchmark data. In the experiment, the network size (nodes) was considered as 1000 with minimum community size 150 and maximum community size 250. The number of communities generated was seven or eight.

In Fig. 12a, variation with η is shown for $O_n = 0.15$, whereas in Fig. 12b, the variation of O_n corresponds to $\eta = 0.4$. As expected, NFMI decreases in all the cases when η increases. For $\eta > 0.3$, FRC-FGSN shows prominent improvement over the other methods. FRC-FGSN performs superior for O_n ranging from 0.2 to 0.4 and second best for $O_n < 0.2$. Therefore, FRC-FGSN has a strong promise for detecting various overlapping communities in Big data analysis. For other relevant references in this area of research, one may cite the studies in [53–58].

9 Fuzzy–Rough Granular Neural Networks

9.1 Principle

Fuzzy sets are used to generate input vector in terms of fuzzy granules (low, medium and high) and output vector in terms of membership value and zeros.

Fuzzy–rough sets are used to extract the domain knowledge from the data, which is then incorporated as the initial parameters (connection weights) of an artificial neural network for its formation.

The resulting network, involving an integration of artificial neural networks, fuzzy set theoretic granules and fuzzy–rough set theoretic knowledge encoding, is termed as fuzzy rough granular neural networks (FRGNN). This network, based on fuzzy–rough knowledge encoding, is found to be superior in handling uncertainty, reducing complexity and in improving the performance as compared to that based on rough set-based encoding.

9.2 Lower Approximation and Knowledge Encoding

Using the concept of fuzzy–rough sets, here each class is characterized by a pair of fuzzy lower and fuzzy upper approximate regions. That is, the memberships of a pattern to lower and upper approximate regions of different classes are fuzzy. Since the input patterns are represented in terms of fuzzy granules *low, medium* and high, the lower and upper memberships have three components (corresponding to low, medium, high of conditional attributes/input features). Each of the components is determined from the similarity matrix of patterns, defined over that component, with respect to fuzzy decision attributes/class labels.

In other words, one can compute the membership values of all the patterns to lower approx. regions of their own classes based on the fuzzy similarity relation corresponding to a conditional attribute *low, medium* or *high*. Therefore, for a class $c1$ with $n1$ patterns, say, finding the degree of membership of every pattern to lower approx. region of that class corresponding to *l.m.h* components would result in $3n1$ lower approx. values for that class. These $3n1$ membership values together constitute what is called the "fuzzy information granule" of class $c1$.

These membership values with respect to lower approx. regions are used to compute the dependency factors of *l*th, *m*th and *h*th components of a feature, which are then used as the corresponding initial network parameters, as they reflect the domain knowledge. This is what is called knowledge encoding.

Note that for c-class problem, we split the training samples in c parts and construct c decision tables based on them. Accordingly, we configure c initial networks which are then integrated and trained with back-propagation based on the gradient decent method.

In a nut shell, the principle of forming FRGNN is as follows:

- Detect lower approximation of classes.
- Find class information granules with lower approx. membership values—called *Knowledge*.
- Form basic networks of the classes by encoding this information.
- Grow these networks with samples of "upper–lower" regions to generate the final network.

That is, the initial network formation is based on pure samples of lower approximate regions (core regions) about which there is no doubt about their class belonging. These networks are then evolved or grown by the samples of upper-lower approximate regions, which are ambiguous (not pure).

9.3 Performance Analysis

Figure 13 shows the performance of the FRGNN [59] with three different weights for a speech data set for the following three conditions:

(a) FRGNN with weights within [0, 1].
(b) FRGNN with weights in crisp case.
(c) FRGNN with weights in fuzzy case.

The speech data contains six vowel classes of Indian Telugu language. It has 871 samples with three features

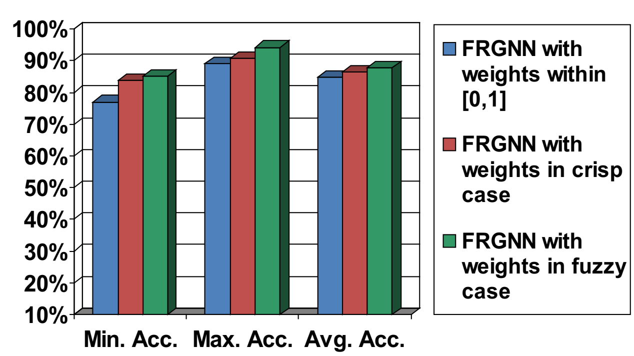

Fig. 13 Performance of FRGNN in terms of maximum, minimum and average accuracy

obtained from the utterances of words in CVC (consonant–vowel-consonant) format by three male speakers in the age group of 30–35 years. Features are the first, second and third format frequencies. Six vowel classes have significant amount of overlapping. The results shown correspond to tenfold cross-validation with ninefolds for training and onefold for testing. Maximum, minimum and average accuracy values, computed over ten such results, are plotted. As expected, the results corresponding to the case "FRGNN with weights in fuzzy case" are superior in terms of any of the aforesaid accuracy parameters. That is, the consideration of fuzzy membership (instead of crisp) in lower and upper approximate regions seems to be appropriate in modeling the uncertainty arising from overlapping characters of the classes. Assigning random weights in [0, 1] provides worst performance.

Figure 14 depicts the comparison between FRGNN and rough–fuzzy MLP [60] using the aforesaid accuracy measures. In FRGNN, fuzzy–rough sets are integrated with a fuzzy MLP (fuzzy multi-layer perceptron), whereas in the latter case it is rough sets which are integrated with a fuzzy MLP. From the results, it is clear that the knowledge encoding based on fuzzy–rough set is better than that of rough set, thereby providing superior performance. This is also verified in Fig. 15 which shows the variation of error rate with the number of epochs during training of the network. In both cases, as expected, the error rate decreases with epoch. However, the amount of error has always been less for FRGNN.

The aforesaid design is shown for pattern classification in supervised mode. The similar concept can be used in forming SOM (self-organizing map) [61, 62], and for feature selection in unsupervised mode [63]. The recent research monograph [64] may be referred for detailed discussion. One may refer to [65–68] for other approaches

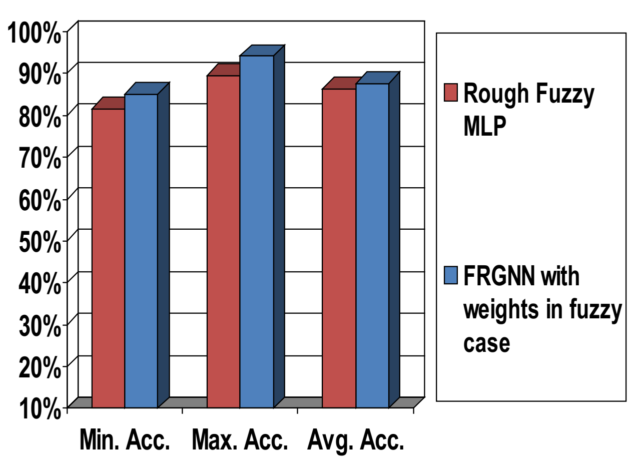

Fig. 14 Comparison between FRGNN and rough–fuzzy MLP in terms of maximum, minimum and average accuracy

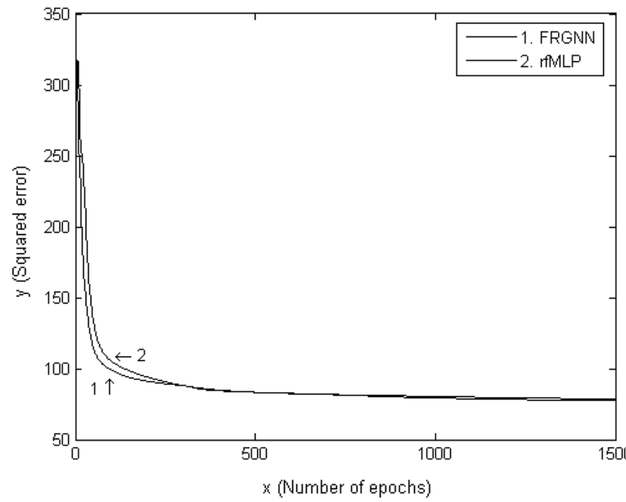

Fig. 15 Comparison between FRGNN and rough–fuzzy MLP in terms of error during training

of designing granular networks and defining rough fuzzy measures.

10 Conclusions and Discussion

The article addresses the significance of GrC in several mining applications, such as in video analysis, bioinformatics and online social networks which are considered to have all the aforesaid characteristics of Big data. After describing the basic components of GrC and the relevance of fuzzy set and rough sets therein, it explains some salient features such as generalized rough sets and rough information granules which are significant for uncertainty modeling and dimensionality reduction in mining large data sets. Formation of information granules and lower approximation (the sense of definite belonging) for the aforesaid mining applications is illustrated. These are seen to be effective, for example, in extracting the object model from videos for unsupervised tracking, estimating the definite and doubtful regions while ranking miRNAs for cancer detection, detecting fuzzy–rough communities in social networks, and in encoding domain knowledge for designing granular neural network models. The resulting systems exhibit superior performance as compared to other related models.

The technologies and models described here basically provide new machine learning modules. Although some specific applications are demonstrated, they can be applied to other real-life problems in data science [69]. As the investigations suggest, the BDA (Big data analytics) can be enriched by these modules from the points of uncertainty management and granular mining points of view. However, one may also look into the scalability and computational aspects, in addition.

Besides, the aforesaid study has the immediate scope of research in the following areas, among others:

- CTP (Computational Theory of Perception) [70]—where the computation is performed based on perception, rather than the measurement. Because of the limited resolution capability of human brain, the boundaries of perceptions are imprecise (fuzzy) and the attribute values that they can take are granules. These *fuzzy (f)-granularity* characteristics of CTP can be modeled using the aforesaid *fuzzy–rough* computing concept.

For example, consider the concept of perception granule of a statement—"The flower is beautiful" in natural language processing (NLP). Here the perception granule is: {subject, predicate, belief} = {description of flower, beautiful, how strongly one *Believes* to be beautiful}.

There are two levels of precisiation in this perception granule. Level 1 precisiation is associated with the term "beautiful." It means, how much (to what extent) beautiful, e.g., very beautiful, more or less beautiful. Level 2 precisiation is associated with the words "how strongly one *Believes* to be beautiful." Here Belief means subjective probability that the flower is beautiful. Belief could be "adverb," or "adjective" or "adverbial phrase" [71].

Zadeh's *z*-numbers [72] provide quantification of precisiation of such perception granules. And *Z**-numbers [73], which are augmented *z*-numbers, characterize the subjectivity (e.g., context/semantics), in addition, for abstraction of active thoughts for such a sentence. A model for *Z**-number-based thought processing is recently reported [74] in connection with machine–mind development.

- Natural Computing—this is an emerging area where the processes and technologies involved are basically abstracted from natural phenomena. 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. Fuzzy* granulation or *f*-granulation is inherent in human thinking and reasoning process and plays an essential role in human cognition.
- Deep Learning—this is a costly process in terms of time and resource requirement. Here the concept of granulation can be incorporated, for example, at the input convolution layer so that the scanning for the purpose of filtering is made only over granules, instead of each pixel, thereby reducing the computation time significantly. However, in designing such a *Granulated deep learning* framework, one should keep in mind a balanced trade-off between speed and accuracy [75].

Acknowledgements The author acknowledges the DAE Raja Ramanna Fellowship and Sir J.C. Bose Fellowship of the Govt. of India. A part of the work was done while he held an INSA Distinguished Professorship Chair.

References

1. Pal SK, Meher SK (2013) Natural computing: a problem solving paradigm with granular information processing. Appl Soft Comput 13(9):3944–3955
2. Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst 90:111–127
3. Pawlak Z (1982) Rough sets. Int J Comput Inf Sci 11:341–356
4. Pawlak Z (1991) Rough sets: theoretical aspects of reasoning about data. Kluwer Academic, Dordrecht
5. Zadeh LA (1996) Fuzzy logic = computing with words. IEEE Trans Fuzzy Syst 4(2):103–111
6. Pedrycz W (2001) Granular computing: an emerging paradigm. Physica-Verlag, Heidelberg
7. Polkowski L, Skowron A (1998) Towards adaptive calculus of granules. In: Proceedings of the 7th IEEE international conference on fuzzy system, Anchorage, AK, USA, May 1998, pp 111–116
8. Yao JT, Vasilakos AV, Pedrycz W (2013) Granular computing: perspectives and challenges. IEEE Trans Cybern 43(6):1977–1989
9. Komorouski J, Pawlak Z, Polkowski L, Skowron A (1999) Rough sets: a tutorial. In: Pal SK, Skowron A (eds) Rough fuzzy hybridization: a new trend in decision making. Springer, Singapore, pp 3–98
10. Pawlak Z, Skowron A (2007) Rudiments of rough sets. Inf Sci 177:3–27
11. Sen D, Pal SK (2009) Generalized rough sets, entropy and image ambiguity measures. IEEE Trans Syst Man Cybern Part B 39(1):117–128
12. Pal SK (2012) Granular mining and rough–fuzzy pattern recognition: a way to natural computation, (Feature Article). IEEE Intell Inf Bull 13(1):3–13
13. Pal SK, Mitra P (2004) Case generation using rough sets with fuzzy discretization. IEEE Trans Knowl Data Eng 16(3):292–300
14. Qian Y, Lian J, Yao Y, Dang C (2010) MGRS: a multi-granulation rough set. Inf Sci 180:949–970
15. Pal SK, Meher SK, Dutta S (2012) Class-dependent rough–fuzzy granular space, dispersion index and classification. Pattern Recognit 45(7):2690–2707
16. Zhu W, Wang FY (2007) On three types of covering-based rough sets. IEEE Trans Knowl Data Eng 19(8):1649–1667
17. Maggio E, Cavallaro A (2010) Video tracking: theory and practice. Wiley, New York
18. Pal SK, Petrosino A, Maddalena L (eds) (2012) Handbook on soft computing for video surveillance. CRC Press, Boca Raton
19. Pal SK, Bhunia Chakraborty D (2017) Granular flow graph, adaptive rule generation and tracking. IEEE Trans Cybern 47(12):4096–4107
20. Pawlak Z (2005) Flow graphs and data mining. Springer, Heidelberg
21. Bhunia Chakraborty D, Pal SK (2016) Neighborhood granules and rough rule-base in tracking. Natural Computing (special issue

on pattern recognition and mining), Springer, vol. 15, no. 3, pp 359–370
22. Chakraborty D, Uma Shankar B, Pal SK (2013) Granulation, rough entropy and spatiotemporal moving object detection. Appl Soft Comput 13(9):4001–4009
23. Bhunia Chakraborty D, Pal SK (2018) Neighborhood rough filter and intuitionistic entropy in unsupervised tracking. IEEE Trans Fuzzy Syst 26:2188–2200. https://doi.org/10.1109/tfuzz.2017.2768322
24. Yilmaz A, Javed O, Shah M (2006) Object tracking: a survey. ACM Comput Surv 38:1264–1291
25. Comaniciu D, Ramesh V, Meer P (2003) Kernel-based object tracking. IEEE Trans Pattern Anal Mach Intell 25:564–577
26. Cucchiara R, Grana C, Piccardi M, Prati A (2003) Detecting moving objects, ghosts and shadows in video streams. IEEE Trans Pattern Anal Mach Intell 25:1337–1342
27. Fang H, Jiang J, Feng Y (2006) A fuzzy logic approach for detection of video shot boundaries. Pattern Recognit 39:2092–2100
28. Pan P, Schonfeld D (2011) Video tracking based on sequential particle filtering on graphs. IEEE Trans Image Process 20(6):1641–1651
29. Shen C, Kim J, Wang H (2010) Generalized kernel-based visual tracking. IEEE Trans Circuits Syst Video Technol 20:119–130
30. Maddalena L, Petrosino A, Ferone A (2008) Object motion detection and tracking by an artificial intelligence approach. Int J Patt Recognit Artif Intell 22:915–928
31. Dai S, Ren W, Gu F, Huang H, Chang S (2008) Implementation of robot visual tracking system based on rough set theory. In: Proceedings of the fifth international conference on fuzzy systems and knowledge discovery (FSKD 2008), IEEE Computer Society, vol 2, pp 155–160
32. Zhang K, Zhang L, Yang M-H (2014) Fast compressive tracking. IEEE Trans Pattern Anal Mach Intell 36:2002–2015
33. Huang C-M, Fu L-C (2011) Multitarget visual tracking based effective surveillance with cooperation of multiple active cameras. IEEE Trans Syst Man Cybern B Cybern 41L:234–247
34. Nawaz T, Poiesi F, Cavallaro A (2014) Measures of effective video tracking. IEEE Trans Image Process 23:376–388
35. Pal JK, Ray SS, Chow SB, Pal SK (2018) Fuzzy-rough entropy measure and histogram based patient selection for miRNA ranking in cancer. IEEE/ACM Trans Comput Biol Bioinf 15(2):659–672
36. Pal JK, Ray SS, Pal SK (2016) Identifying relevant group of miRNAs in cancer using fuzzy mutual information". Med Biol Eng Comput 54(4):701–710
37. Pal JK, Ray SS, Pal SK (2017) Fuzzy mutual information based grouping and new fitness function for PSO in selection of miRNAs in cancer. Comput Biol Med 89:540–548
38. Maji P, Pal SK (2010) Fuzzy-rough sets for information measures and selection of relevant genes from microarray data. IEEE Trans Syst Man Cybern Part B Cybern 40(3):741–752
39. Yu L, Han Y, Berens ME (2012) Stable gene selection from microarray data via sample weighting. IEEE/ACM Trans Comput Biol Bioinf 9:262–272
40. Sehhati M, Mehridehnavi S, Rabbani H, Pourhossien M (2015) Stable gene signature selection for prediction of breast cancer recurrence using joint mutual information. IEEE/ACM Trans Comput Biol Bioinf 12:1440–1447
41. Guyon J, Weston S, Barnhill V (2002) Vapnik, Gene selection for cancer classification using support vector machines. Mach Learn 46:389–422
42. Peng H, Long F, Ding C (2005) Feature selection based on mutual information: criteria of max-dependency, max-relevance and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27:1226–1238
43. Mundra PA, Rajapakse JC (2010) SVM-RFE with MRMR filter for gene selection. IEEE Trans Nanobiosci 9:31–37
44. Mitra P, Murthy CA, Pal SK (2012) Unsupervised feature selection using feature similarity. IEEE Trans. Pattern Anal Mach Intell 24:301–312
45. Arndt GM, Dossey L, Cullen LM, Lai A, Druker R, Eisbacher M, Zhang C, Tran N, Fan H, Retzlaff K, Bittner A, Raponi M (2009) Characterization of global microRNA expression reveals oncogenic potential of mir-145 in metastatic colorectal cancer. BMC Cancer 9:1–17
46. Leidinger P et al (2010) High-throughput miRNA profiling of human melanoma blood samples. BMC Cancer 10:1–11
47. Kundu S (2016) Granular model for social networks, target selection and fuzzy-rough community detection, Ph.D. Dissertation, Jadavpur University, Kolkata, India
48. Kundu S, Pal SK (2015) FGSN: fuzzy granular social networks—model and applications. Inf Sci 314:100–117
49. Kundu S, Pal SK (2015) Fuzzy-rough community in social networks. Pattern Recognit Lett 67(2):145–152
50. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci USA 99(12):7821–7826
51. Newman MEJ (2004) Fast algorithm for detecting community structure in networks. Phys Rev E 69(6):066133
52. Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435(7043):814–818
53. Boorman SA, White HC (1976) Social structure from multiple networks. II. Role structures social structure from multiple networks. Am J Sociol 81:1384–1446
54. Davis GB, Carley KM (2008) Clearing the FOG: fuzzy, overlapping groups for social networks. Soc Netw 30:201–212
55. Newman M, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69:1–15
56. Chattopadhyay S, Murthy CA, Pal SK (2014) Fitting truncated geometric distributions in large scale real world networks. Theoret Comput Sci 551:22–28
57. Malliarosa FD, Vazirgiannis M (2013) Clustering and community detection in directed networks: a survey. Phys Rep 533:95–142
58. Weiss RS, Jacobson E (1955) A method for the analysis of the structure of complex organizations. Am Sociol Assoc 20:661–668
59. Ganivada A, Dutta S, Pal SK (2011) Fuzzy rough granular neural networks, fuzzy granules and classification. Theor Comput Sci Part C 412(42):5834–5853
60. Banerjee M, Mitra S, Pal SK (1998) Rough fuzzy MLP: knowledge encoding and classification. IEEE Trans Neural Netw 9(6):1203–1216
61. Ganivada A, Ray SS, Pal SK (2012) Fuzzy rough granular self-organizing map and fuzzy rough entropy. Theoret Comput Sci 466:37–63
62. Ray SS, Ganivada A, Pal SK (2016) A granular self-organizing map for clustering and gene selection in microarray data. IEEE Trans Neural Netw Learn Syst 27(9):1890–1906
63. Ganivada A, Ray SS, Pal SK (2013) Fuzzy rough sets, and a granular neural network for unsupervised feature selection. Neural Netw 48:91–108
64. Pal SK, Ray SS, Ganivada A (2017) Granular neural networks, pattern recognition and bioinformatics. Springer, Berlin
65. Zhang YQ, Jin B, Tang Y (2008) Granular neural networks with evolutionary interval learning. IEEE Trans Fuzzy Syst 16:309–319
66. Pal SK, Dasgupta B, Mitra P (2004) Rough self organizing map. Appl Intel 21:289–299
67. Yeung DS, Chen D, Tsang ECC, Lee JWT, Xizhao W (2005) On the generalization of fuzzy rough sets. IEEE Trans Fuzzy Syst 13:343–361

68. Banerjee M, Pal SK (1996) Roughness of a fuzzy set. Inf Sci 93:235–246
69. Pal SK, Meher SK, Skowron A (2015) Data science, big data and granular mining. Pattern Recognit Lett 67(2):109–112
70. Zadeh LA (2001) A new direction in AI: toward a computational theory of perceptions. AI Magazine 22:73–84
71. Pal SK, Banerjee R (2013) Context granularization and subjective-information quantification. Theoret Comput Sci 448:2–14
72. Zadeh LA (2011) A note on *Z*-numbers. Inf Sci 18(14):2923–2932
73. Banerjee R, Pal SK (2015) *Z**-numbers: augmented *Z*-numbers for machine-subjectivity representation. Inf Sci 323:143–178
74. Banerjee R, Pal SK (2017) A computational model for the endogenous arousal of thoughts through *Z**-numbers. Inf Sci 405:227–258
75. Bhoumik D (2018) Granulated deep learning: application in video tracking and object recognition, M. Tech. (CSE) Dissertation, Department of Computer Science and Engineering, University of Calcutta, India