Intelligent Decision Technologies 10 (2016) 115–128
DOI 10.3233/IDT-150243
IOS Press

Soft granular computing based classification using hybrid fuzzy-KNN-SVM

Mrutyunjaya Panda[a,*], Ajith Abraham[b,c] and B.K. Tripathy[d]
[a]*Department of Computer Science, Utkal University, Vani Vihar, Odisha, India*
[b]*Machine Intelligence Research Labs (MIR Labs), Auburn, WA, USA*
[c]*IT4Innovations, Center of Excellence VSB, Technical University of Ostrava, Ostrava, Poruba, Czech Republic*
[d]*School of Computing Science and Engineering, VIT University, Vellore, Tamilnadu, India*

Abstract. This paper aims at providing the concept of information granulation in Granular computing based pattern classification that is used to deal with incomplete, unreliable, uncertain knowledge from the view of a dataset. Data Discretization provides us the granules which further can be used to classify the instances. We use Equal width and Equal frequency Discretization as unsupervised ones; Fayyad-Irani's Minimum description length and Kononenko's supervised discretization approaches along with Fuzzy logic, neural network, Support vector machine and their hybrids to develop an efficient granular information processing paradigm. The experimental results show the effectiveness of our approach. We use benchmark datasets in UCI Machine Learning Repository in order to verify the performance of granular computing based approach in comparison with other existing approaches. Finally, we perform statistical significance test for confirming validity of the results obtained.

Keywords: Granular computing, discretization, supervised model, unsupervised model, hybrid model, statistical significance

1. Introduction

Granular computing was first proposed by Lin [48] and has become a very important tool in data mining since then. The term granule is originated from Latin granum, i.e., grain, to denote a small compact particle in physics and in the natural world. The taxonomy of granules in computing can be classified as: data granule, information granule, concept granule, computing granule, cognitive granule, and system granule [7,28,56,58]. Granular computing seems to cover any theories, methodologies, techniques, and tools that make use of granules in pattern recognition. While a subset of the universe is called a granule in granular computing in terms of: subsets, classes, and clusters of a universe. It deals with the characterization of a concept by a unit of thoughts consisting of two parts, the intension and extension of the concept [59].

*Corresponding author: Mrutyunjaya Panda, Department of Computer Science, Utkal University, Vani Vihar, Bhubaneswar-4, Odisha, India. E-mail: mrutyunjaya74@gmail.om.

Granular computing is a general computing methodology using granules to establish an efficient computational model for complex applications [1]. The theory of fuzzy information granulation (TFIG) is inspired by the ways in which humans granulate information and reason with it [29]. TFIG builds on the existing machinery of fuzzy information granulation in fuzzy logic but takes it to a significantly higher level of generality, consolidates its foundations and suggests new directions [29,30]. Although extensive work has been done on granular computing, it still attracts attentions of researchers in the field of data mining. In this paper, we will attempt to provide such a Label of theories, methodologies, techniques, and tools that make use of granules in the process of problem solving in pattern recognition.

The motivation behind using granular computing for our proposed information granulation with subsequent successful classification data mining rests on: (i) information granulation is very essential to human problem solving, and hence has a very significant impact on the design and implementation of intelligent systems, from

ISSN 1872-4981/16/$35.00

theoretical aspects and (ii) then, the necessity of information granulation and simplicity derived from information granulation in problem solving in view of practice. From these, we understand that, procedure for information granulations become inevitable when a problem involves incomplete, uncertain, or vague information and it becomes difficult to differentiate distinct elements. A typical example is the theory of rough sets [63]. The lack of information may only allow us to define granules rather than individuals. In some situations, even though detailed information are available, it still can use granular computing for an efficient and practical solution. This may reduce cost as the acquisition of precise information is too costly. The main reason behind using fuzzy granular computing lies in dealing with imprecision, uncertainty and partial truth to achieve tractability, robustness, low solution cost and better practical solution to a problem under investigation.

The granular computing [10] is a branch of soft computing, considered to be a new concept and computing paradigm of information processing, mimic the way of human's thinking. It uses a low-cost approximate solution instead of the precise solution for developing a robust and efficient intelligent systems for better understanding. Further, we can envisage information granules are pivotal to a way in which experimental datasets are perceived and interpreted by humans, where one needs to summarize the large datasets and forms a very limited vocabulary of information granules that are easily comprehended and essential dependencies among individual variables arising at some specific level of information granularity is understood.

Granular Computing has become indispensable in machine learning in several fundamental ways:

– Granular Computing helps us to build heterogeneous models in processing of information granules using fuzzy sets.
– Granular Computing considers the notion of variable granularity whose range could cover detailed numeric entities that are very abstract and general information granules.

1.1. Advantages of granular computing (GrC)

Individually, Fuzzy set theory, rough set theory and quotient space theory have their own advantages and disadvantages, but they constitute as the three main subjects of granular computing in unison for dealing with complex and ambiguous information more easily. Miao et al. [11] summed up the three models' advantages and proposed a new model: using rough set to construct a collection of researching objects (domain size), using fuzzy sets to make sure the semantic interpretation of granularity, using quotient space theory to construct a hierarchical granular structure when they summarize the granular computing in order to solve the problems conveniently. The proposed models are not only suitable for solving complex information systems, but also has good adaptability for the uncertainty in the data calculation followed by low cost, robustness, feasibility, and imitation human intelligence to name some of its advantages.

1.2. Dis-advantages of GrC

Apart from its numerous advantages, granular computing also has some disadvantages. Diversity of the granularity leads to a variety of different representations and interpretations increases the difficulty of understanding. This may be avoided by specifying the size of the granularity for achieving an efficient granular based model.

In this paper, we address some of the basic issues of granular computing such as: construction of granules and computation with granules using classification data mining. Construction of granules as a first part of information granulation makes us understand why two objects are put into the same granule in terms of their similarity, proximity, or functionality etc. [29]. Furthermore, information granulation depends on the available knowledge, where it is necessary to study criteria for deciding if two elements should be put into the same granule, based on information at hand. Novel Algorithms need to be developed for constructing granules efficiently considering how to put two objects into the same granule. Granular computation with classification can be similarly studied from both the semantic and algorithmic perspectives [50,60].

To summarize, we may consider the following points:

– Although detailed information is available, it may be sufficient to use granules in order to have an efficient and practical solution.
– Very precise solutions may not be required at all in many practical problems.
– Acquisition of precise information is too costly.
– Instead of searching for optimal solution one may search for good approximate solutions.
– Studies of granular computing are only complementary to vigorous investigations on precise and non-granular computational approaches.

1.3. Research gaps

It is observed that previous works on granular support vector machine (GSVM) models, often referred to as traditional ones, are granulated on the original space and trained on the kernel space can improve the learning efficiency, but at the cost of some losses in performance [53]. This is because (i) after granulation, different data distribution exists between original space and kernel space and (ii) using granulation before training and then use some sample granules from whole for training that may lead to errors in data distributions.

Considering the above, a novel granular support vector machine model in combination with supervised and unsupervised discretization, are proposed for fine granulation for efficient data mapping between original space and kernel space. Then, the granules obtained are processed for classification data mining using Fuzzy K-nearest neighbor, Multilayer perceptron etc. that can largely improve the generalization performance with the high accuracy.

The rest of the paper is structured as follows. Some of the related work is discussed in Section 2. Section 3 is concerned with the information granulation using discretization techniques proposed. Subsequently, in Sections 4 and 5, we proceed with the realization of our proposed methodology with the experimental results obtained along with discussions. Finally, Conclusions with future scope of research are covered in Section 6.

2. Related works

Yao and Yao [26] presented a granular computing view to classification problems using ID3 and PRISM algorithm. In this, they state that a granule is a subset of granular computing that deals with the characterization of a concept by a unit of thoughts consisting of the intension and extension of the concept. The role of Granular Computing as a fundamental environment supporting the development of intelligent systems is discussed by Pedrycyz [51]. The authors presented various conceptual alternatives (such as interval analysis, fuzzy sets, and rough sets) and provided an general overview of Granular Computing using fuzzy logic and fuzzy sets to construct a consistent processing necessary for operating on information granules. Finally, the authors conclude on how logic neurons can contribute to high functional transparency of granular processing, help capture prior domain knowledge and give rise to a diversity of the resulting models. Shifei et al. [13] discussed about principle of granularity in clustering using k-means, leader and spectral clustering algorithm with their pros and cons. The authors then evaluated their model using rough set, fuzzy sets and quotient space theories for its effectiveness.

A critical analysis of the granular computing paradigm with system algebra has been outlined for granular computing and granule-based systems modeling by Wang et al. [57]. Mooe and Inoue [62] reported the effectiveness of value granulation (such as discretization and quantization) in Machine Learning from aspects of complexity reduction, learning capability, and intelligent system development and then demonstrated the effectiveness of value granulation in machine learning in terms of low computation and size reduction, used SVM for demonstrating the effectiveness of the value granulation but could not succeeded. Sikdar [21] presented Rough set theory based granular computing approach with its variable precision extension. The memberships function used in variable precision extension of rough set generalize the lower and upper approximations for handling decision analysis problems to incorporate qualitative knowledge in the knowledge induction process.

Li and Ding [17] opines that Granular neural networks (GNNs) based on Granular Computing (GrC) and artificial neural network can be able to deal with all kinds of granular information of the real world. They provided an introduction to the basic model of GrC: word calculation model based on fuzzy sets theory, rough sets model, and granular computing model based on quotient space theory and then analyses the ensemble-based methods of GNNs. Chen et al. [54] have demonstrated the application of granular computing model in outlier detection. They proposed a GrC-based method for effective outlier detection using three publicly available databases. Zhang and Cheng [33] investigates new approaches to solve the pattern classification problems using granularity computing of quotient space theory. Here, they use granular computing to solve the classification problems with incomplete information system. Liu et al. [18] proposed Granular computing based classification algorithms taking distance between two granules from the view of set. They use different forms of the granule as: hyperdiamond, hypersphere, hypercube, and hyperbox in order to form the granular computing classification algorithms based on distance measures (DGrC). The benchmark datasets in UCI Machine Learning Repository are used by the authors to verify the performance of DGrC, with improved the testing accuracies.

Panoutsos and Mahfouf [15] presented a new systematic modelling approach using Granular Computing (GrC) and Neuro-fuzzy modelling, where GrC is used to extract relational information and data characteristics from the used dataset. The extracted knowledge is then translated into a linguistic rule-base of a fuzzy system followed by a Neuro-fuzzy modelling structure that has been tested against real industrial data of high dimensionality and complex nature. The authors claim that the resulting models achieved better performances than pure fuzzy models. Bargiela and Pedrycyz [2] proposed a structured combination of algorithm and non-algorithm information processing for developing a Granular computing paradigm.

Ganivada et al. [4] introduced a fuzzy rough granular neural network (FRGNN) model based on the multilayer perceptron using a back-propagation algorithm for the fuzzy classification of patterns. The effectiveness of the proposed FRGNN is demonstrated on several real-life data sets with two special types of granular computation such as: low, medium, and high fuzzy granules and the other has classes of granulation structures induced by a set of fuzzy equivalence granules based on a fuzzy similarity relation. The authors concluded that FRGNN to be superior to a rough fuzzy MLP which uses rough sets, rather than fuzzy rough sets, for knowledge encoding for vowel data.

Cimino et al. [34] dealt with information granules, namely interval-valued data and propose a multilayer perceptron (MLP) to model interval-valued input-output mappings followed by using a genetic algorithm to fit data with different levels of granularity. The results of the proposed MLP are illustrated by means of its application to both synthetic and real world datasets. Tang et al. [55] used microarray gene expression dataset with Fuzzy Granular Support Vector Machine – Recursive Feature Elimination algorithm (FGSVM-RFE) as a hybrid algorithm of statistical learning, fuzzy clustering, and granular computing. The authors pointed out that the results obtained on three public datasets demonstrate that the FGSVM-RFE outperforms state-of-the-art approaches.

Pal [45] introduced Rough-fuzzy granular approach in natural computing framework with the concept of rough set theoretic knowledge encoding and the role f-granulation for its improvement. Even though new machine learning modules are discussed, they should be applied to other real life problems application for the understanding of its efficacy. Zhu and Hu [37] addressed the granular computing model using neighbourhood rough sets to deal with complex tasks of classification learning. Here, the author highlighted the importance of base classifiers for training in different granular spaces in a combined system. They conclude that the proposed methods are effective and the derived models produce competent performances compared with other classical techniques in terms of squared loss for training weights of different granularity.

Pal and Meher [46] provided an overview of the significance of natural computing with respect to the granulation-based information processing models, such as neural networks, fuzzy sets and rough sets, and their hybridization. Ganovada et al. [5] proposed a granular neural network for identifying salient features of data, based on the concepts of fuzzy set and a newly defined fuzzy rough set. The author assigns each feature of the data between 0 and 1 and used to develop granulation structures by a user defined α-value. The input vector and the target value of the network are defined using granulation structures, based on the concept of fuzzy sets. The authors pointed out that the results of FRGNN are found to be statistically more significant than related methods in 28 instances of 40 instances, i.e., 70% of instances, using the paired t-test, but takes higher computational time than the related methods for feature selection tasks.

Pal et al. [47] addressed the problem of image object extraction in the framework of rough sets and granular computing with rough entropy of image based on the concept of image granules. Yao [27] presented an overview of granular computing research in the past ten years using two popular databases, ISI's Web of Science and IEEE Digital Library where current status, the trends and the future direction of granular computing and identify prolific authors, impact authors, and the most impact papers in the past decade are studied. Zadeh [31] opined the importance of soft computing, fuzzy logic, granular computing and computing with words that are essential to the conception, design and utilization of information/intelligent systems. Pedrycyz [52] stressed on the role of associations of fuzzy sets, formalism of fuzzy sets (and fuzzy clustering), and investigated in depth about information granules (sets, rough sets, shadowed sets and others). However, they did not report about the classification accuracy.

Li et al. [12] proposed a new granular computing based incremental recursive reduction algorithm in information system claiming improved computation accuracy and decrease in cost.

Hu and Guan [25] proposed a new emotional reasoning algorithm, and presents an emotional agent model based on granular computing. They perform the simulation in hospital scenario presenting an emotional agent for patient assistance where results show its goodness to transact simple emotions.

Ding and Qi [41] pointed out the research mostly concentrated on data granulation and its simplicity in hierarchical space with no effective standards, causing unstable generalization ability of granular support vector machine (GSVM). They suggested the integration of Granular computing and SVM needs further research.

Considering the constraints of time and memory on learning performance of Support Vector Machine (SVM) for large number of samples, Huang et al. [16] proposed a novel algorithm called Granular Support Vector Machine based on Mixed Kernel Function (GSVM-MKF). Here, a new granular support vector machine learning model is constructed with support vector particles. Next, a new GSVM based on mixed kernel function (GSVM-MKF) is proposed by the authors for the strong generalization ability and strong learning ability of mixed kernel function, stating on how to choose threshold value effectively as a further course of study.

Singh et al. [39] presented a novel formulation of multi-class support vector machine by integrating the concepts of

Soft labels and granular computing, in order to address the issues of noisy and incorrectly labelled data for efficient data distributions, in a multispectral face recognition application.

Kazem et al. [6] developed a novel hybrid model based on a chaotic firefly algorithm and support vector regression for stock market price forecasting. Their contribution of the proposed algorithm is mainly the integration of chaotic motion with a firefly algorithm as a simple and novel optimization method. Compared with genetic algorithm-based SVR (SVR-GA), chaotic genetic algorithm-based SVR (SVR-CGA), firefly-based SVR (SVR-FA), artificial neural networks (ANNs) and adaptive neuro-fuzzy inference systems (ANFIS), the proposed model performs best based on two error measures, namely mean squared error (MSE) and mean absolute percent error (MAPE).

Wan et al. [8] implemented a new text document classifier by integrating the K-nearest neighbor (KNN) classification approach with the support vector machine (SVM) training algorithm. The proposed Nearest Neighbor-Support Vector Machine hybrid classification approach is coined as SVM-NN which avoids a major problem of the KNN in determining the appropriate value for parameter K in order to guarantee high classification effectiveness. By considering several benchmark text datasets for their experiments, it is shown that the classification accuracy of the SVM-NN approach has low impact on the value of parameter, as compared to the conventional KNN classification model.

Lin et al. [9] proposed a support-vector-based fuzzy neural network (SVFNN) to minimize the training and testing error for better performance. They developed a learning algorithm consisting of three learning phases is to construct the SVFNN in which the fuzzy rules and membership functions are automatically determined by the clustering principle. To investigate the effectiveness of the proposed SVFNN classification, they applied the corresponding model to various datasets from the UCI Repository and Statlog collection. Experimental results show that the proposed SVFNN for pattern classification can achieve good classification performance with drastically reduced number of fuzzy kernel functions.

We can summarize the section as follows: There are many articles that shows how to perform better information granulation process to achieve high predictive accuracy apart from time needed for discretization and learning, with a simple understanding for numerous applications. However, in this paper, focus is given to apply various supervised and unsupervised discretization methods and discussed various classifiers that can be categorized efficiently, for classification data mining with 7 datasets with and without missing attribute values. Finally, statistical significance test shall be carried out to understand whether one method can significantly outperform others on a given problem.

3. Construction of proposed information granulation approach using discretization

In conventional Granular Computing, there are three types of information granulation present. First, Value Granulation that corresponds to discretization and quantization. Second, Variable Granulation corresponding to clustering, aggregation and transformation and finally, Concept Granulation corresponding to component analysis.

As, value granulation is the simplest as it can be achieved by discretization and quantization, the following procedure is to be performed:

1. Domain X_n is granulated by values such that each dimension is divided into uniform segments $(1 \leqslant i \leqslant n)$.

2. Non-empty segments are considered as information granules. Such information granules are crisp sets and are not overlapped each other.

Data discretization is defined as one of the way to reduce data used to change the original continuous attributes to discrete attributes [38,61]. It aims at creating an appropriate number of intervals for data values thus transforming the continuous data values into the discrete values. The smaller data intervals usually contributed to more accurate predictive model which could cover higher prediction rates into new cases.

Information granulation or data granulation as means of feature reduction using discretization has many advantages since it reduces the content volume and makes it easy to handle lot of information without challenging the system resources. But the technique to discretize or compress data without loss of valuable information is the key challenge. There are many techniques reported in the literature but an algorithms that can suit the given condition needs to be validated by using the transformed data in the developed model for establishing performance improvements.

3.1. Supervised and unsupervised discretization methods

In this section, we present the various Supervised and Unsupervised discretization techniques used in this paper.

In the unsupervised methods, continuous ranges are divided into sub-ranges by the user specified parameter – for say, equal width (specifying range of values), and equal frequency (number of instances in each interval). Further, if no class information is available, unsupervised discretization is the only choice. In supervised discretization methods class information is used to find the proper intervals caused by cut-points. Different methods have been devised to use this class information for finding meaningful intervals in continuous attributes. Supervised discretization can be further characterized as error-based, entropy-based or statistics-based according to whether intervals are selected using metrics based on error on the training data, entropy of the intervals, or some statistical measure [24].

While there are many supervised methods exist in literature, not much work has been done for synthesizing unsupervised methods. This might be due to the fact that discretization has been commonly associated with the classification task. Therefore, work on supervised methods is strongly motivated in those learning tasks where no class information is available. But, when there is no class information available, unsupervised discretization can play a vital role in designing a system.

3.1.1. Unsupervised discretization

Both Equal Width and Equal Frequency discretization are unsupervised discretization methods [24] and have been used because of their simplicity and reasonable effectiveness.

- *Equal Width Discretization* – The simplest unsupervised discretization method, which determines the minimum and maximum values of the discretized attribute and then divides the range into the user-defined number of equal width discrete intervals. There is no “best” number of bins, and different bin sizes can reveal different features of the data. Some theoreticians have attempted to determine an optimal number of bins.
- *Equal Frequency Discretization* – The unsupervised method, which divides the sorted values into k intervals so that each interval contains approximately the same number of training instances. Thus each interval contains n/k (possibly duplicated) adjacent values. k is a user predefined parameter. For our experiments we have chosen k to be 10.

Both EWD and EFW suffer from possible attribute loss on account of the pre-determined value of k.

3.1.2. Supervised discretization methods

Fayyad-Irani Discretization and Kononenko’s Discretization are two supervised discretization methods are considered in our approach.

- *Fayyad-Irani Discretization method* [49] is a supervised hierarchical split method, in which the class information entropy of candidate partitions are used to select boundaries for discretization. Class information entropy is a measure of purity that measures the amount of information which would be needed to specify to which class an instance belongs, considering one big interval containing all known values of a feature. Recursive partitions are then used on this big interval to convert many smaller subintervals until the Minimum Description Length MDL criterion is attained.
  It has been shown that optimal cut points for entropy minimization must lie between examples of different classes. The Fayyad and Irani model [13] uses a supervised hierarchical split method where multiple ranges are created instead of binary ranges to form a tree.

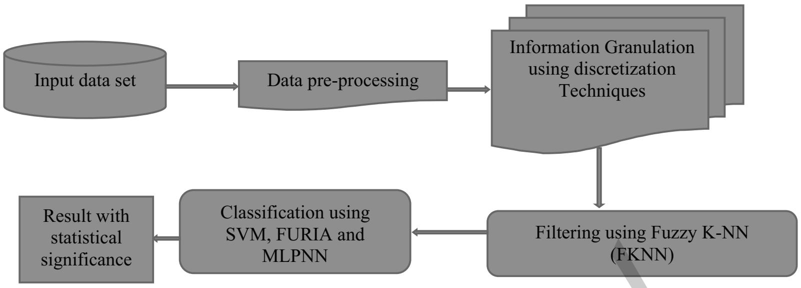

Fig. 1. Proposed methodology.

– *The Kononenko's* discretization [19] uses the Recursive entropy discretization proposed by Fayyad and Irani with a minor alteration by including an adjustment for discretization of multiple attributes. It provides a correction for the bias, the entropy measure has towards an attribute with many values [24,35,40].
The advantages of these data discretization as information granulation provides: regularization as it is less prone to variance in estimation from small fragmented data; amount of data can be greatly reduced because some redundant data can be identified and removed and better performance [20,44].

4. Proposed methodology

As the data extracted from various databases are large in size and high dimensional, we use Granular Computing for more simplicity in our approaches as an underlying scheme of data mining that tackles the fundamentals of the development of intelligent systems.

Our contribution in this paper consists of following steps:

Step 1 is granulation to build a sequence of information granules from the dataset used.

Step 2 uses single best known classifier such as Support vector machine (SVM), Multilayer perceptron neural network (MLPNN) and FURIA (Fuzzy Unordered Rule Induction Algorithm) with supervised and unsupervised discretization for classification data mining.

Step 3 is modelling hybrid Fuzzy-K-nearest neighbour (FKNN) based Support Vector Machines (SVM) in these information granules, as FKNNSVM for improving the classification accuracy in comparison to the results obtained in Step-2.

Step 4 is to obtain abstract information in these granules, suitable for our proposed methodology. A good granulation method to find suitable granules is crucial for modelling a good FKNNSVM classifier. Under this framework, many supervised and unsupervised granulation algorithms including class independence and class dependence criteria are investigated, for classification problems with different characteristics.

Finally, Step 5 uses Wilcoxon's sign rank test and Kruskal-Wallis statistical significance test on accuracy in order to obtain the efficacy of the proposed model.

At first, we present the details about the dataset that are considered for our experimentation, as provided in Table 1.

A general framework of granular computing based classification is presented in Fig. 1. The raw input data is obtained as public dataset from UCI machine repository. The input data set is transformed into desired format using binary to nominal or vice-versa, as a data pre-processing techniques as and when required for data analysis. The discretization process is then employed for obtaining information granules by eliminating data with incomplete value. Consequently, these granules are filtered to get the most significant ones for the classification process using Fuzzy K-nearest neighbour (FKNN); with $k = 10$. Finally, efficient classification is done with the filtered granules using support vector machine (SVM); FURIA (Fuzzy Unordered Rule Induction Algorithm) and MLPNN (Multilayer Perceptron Neural Network). Statistical significance test enables us to understand the triviality of the proposed approach in comparison to others.

The following are details of the proposed techniques with the performance measures to evaluate the model built.

4.1. Accuracy measurement

Accuracy can be defined as the ratio of correct predictions to the total number of instances.

Table 1
Dataset used

Dataset	Total instances	Numeric attributes	Nominal attributes	Missing attributes	Class
Vowel	871	4	–	–	Multi {six overlapping vowel classes (Indian telugu vowel sounds from 1 to 6}
Ecoli	336	7	2	–	multi {cp (cytoplasm), im (inner membrane without signal sequence), pp (perisplasm), imU (inner membrane, uncleavable signal sequence), om (outer membrane), omL (outer membrane lipoprotein), imL (inner membrane lipoprotein), imS (inner membrane, cleavable signal sequence)}
Iris	150	4	1	–	Multi{–Iris Setosa, –Iris Versicolour –Iris Virginica}
Glass	214	11	–	–	Multi {Type of glass: 1:building_windows_float_processed 2:building_windows_non_float_processed 3:vehicle_windows_float_processed 4: vehicle_windows_non_float_processed 5:containers 6:tableware 7: headlamps}
Segmentation	210	19	1	–	multi{brickface, sky, foliage, cement, window, path, grass}
Opt_digit	3823	64	1	–	Multi {class code 0..9}
Post-operative patient data	90	1	8	Yes (attribute 8 has 3 missing values that presents patient's perceived comfort at discharge, measured as an integer between 0 and 20	Multi {I (patient sent to Intensive Care Unit), S (patient prepared to go home) A (patient sent to general hospital floor)}

4.2. Coverage measurement

Coverage can be defined as the amount (%) of a data set for which a classifier makes a prediction. If a classifier does not classify all the instances, then consideration of a set of its confident cases are of paramount importance while making a prediction [36]. The rules are completed if any object belonging to the class is assigned with the coverage of 1, but the rules are correct when both coverage and accuracy equal to 1.

The block diagram of our proposed methodology is shown in Fig. 1.

4.3. Computational time

We use the time taken by a methodology to build a classification model in seconds, as another criteria to evaluate their effectiveness.

4.4. Analysis of errors (RRSE)

In this paper, root relative squared error (RRSE) is used as a performance measures for evaluating the classifier. This can be computed as amount of total squared error made relative to what the error would have been in absolute terms.

4.5. Hybrid methods, our proposal

The K-NN is a supervised learning algorithm where the classification of a query for new instance is done considering the majority of k-Nearest Neighbor category. Here, the following parameters are set for necessary classification: $k = 10$, the window Size is 0 for unlimited number of training instances and 10-fold cross validation, for experimentation.

In Artificial neural networks (ANN), learning the classifier is done by comparing the instances with the known instances at first. The errors resulted from the initial classification is then fed back to the network for further modification in the networks algorithm from the second time onwards till the error is minimized. In this paper, Multilayer Perceptron Neural Network classifier with sigmoid activation function and back propagation as its learning algorithm is used, with the following selected parameters: Hidden Layers is 'a' [the wildcard a = (attributes + classes)/2], Momentum is 0.2, Learning rate is 0.4, Number of Epochs is 500, Random seed for Weights is 0, validation Set Size is 0 and the validation threshold is 20.

Support Vector Machines (SVMs), is popularly used in identifying hyper planes in order to maximize the margins between the two classes, so as to handle multiple continuous and categorical variables. For our experiments we have used SVM classifier- which implements John C. Platt's sequential minimal optimization algorithm for training a support vector classifier using polynomial kernels with the following parameters: complexity parameter c is 1.0, gamma is 1.0, kernel cache size is 250007, Polynomial kernel with exponent 1.0.

This paper proposes a novel fuzzy rule-based classification method called Fuzzy Unordered Rule Induction Algorithm (FURIA), an extension of the rule learner RIPPER. FURIA learns fuzzy rules instead of conventional rules and unordered rule sets instead of rule lists and avoids the default class prediction problem of RIPPER, makes it a most promising one in pattern recognition problem.

The FURIA [22] is based on RIPPER [42] (Pruning to Produce Error Reduction) by William Cohen of AT&T Laboratories which is used here because of its high accuracy in predicate-logics rules generation, the information about the information gain for each attribute would be used for inferring a collection of links each represents the predictive power (in term of information gain) towards the prediction class.

Whereas RIPPER produces hard and inflexible decision boundaries between different classes, FURIA proposes introducing a softer transition between class boundaries through fuzzy rules. It also departs from its predecessor by inducing rules for each class using the one-versus-all method which frees up the classifier from a strict order in which it must be evaluated. Arguably, this increases the comprehensibility as well as the knowledge discovery quality of its rules since they no longer implicitly embody the negated conditions of the previous rules [58]. This however introduces a problem during classification of unseen samples, where a sample may not satisfy any of the generated rules. FURIA addresses this by devising a rule stretching mechanism that generalizes the rules further to ensure a maximum coverage.

Here, we use standard product T-Norm for performing as Fuzzy AND operator; rule stretching operation for uncovered instances; 3-fold operation determines the amount of data used for pruning: One fold is used for pruning, the rest for growing the rules and lastly stopping criteria is used when error rate is more than 50%.

Further, Fuzzy K-Nearest Neighbor (FKNN) classifier is used for its obvious advantages over crisp K-NN as: (i) capability to handle the ambiguous neighbors in classification of current residue and (ii) assignment of fuzzy membership values to residues, as a function of the vector's distance from its K-NN for each class rather than binary decision of their belongingness. Tuning of the parameter is done with the neighboring size (k) and the fuzzy strength (m). Here, we use Euclidean distance as membership function with $k = 10$ and $m = 2$. It should be noted that fuzzy strength in continuous value such as 2.1, 2.2 etc. does not change the result significantly.

Considering the Cons of the KNN with pros of SVM, we propose to use the combination of Nearest Neighbor-Support Vector Machine as SVM-KNN hybrid, which avoids a major problem of the KNN in determining the appropriate value for parameter K in order to guarantee high classification effectiveness [14]. By considering several benchmark datasets for our experimentation, it is shown that the classification accuracy of the SVM-KNN approach has low impact on the value of parameter, as compared to the conventional KNN classification model.

Finally, we propose to hybridize FKNN algorithm with support vector machine (SVM) with information granulation to result granular FKNNSVM as a fast and effective machine learning technique. This is because hybrid methods that combine fuzzy set theory with KNN and SVM, in order to optimize a fuzzy rule base or for searching the space of potential rule bases in a best possible systematic way.

To analyse the differences between the classifiers more closely, we followed the two-step procedure: First, a Wilcoxon sign rank and Kruskal-Wallis statistical significance Test are conducted to test the null hypothesis of classifier performance on predictive accuracy. In case this hypothesis is rejected, which means that the classifiers' performance differs in a statistically significant way.

5. Experimental results and discussion

In this section, following the experimental setup in Fig. 1, in order to assess the validity and performance of the proposed discretization algorithm as information granulation, we have performed experiments on several datasets taken from the UCI repository (*Vowel, ecoli, glass, iris, segmentation, opt_digit and post-operative patient*). These datasets contain a large set

Table 2
Unsupervised discretization for information granulation with accuracy (%)

Data/Algorithm	Equal frequency (EFD)						Equal width (EWD)					
	Vowel	Ecoli	Segmentation	Iris	Glass	Opt-digit	Vowel	Ecoli	Segmentation	Iris	Glass	Opt-digit
Furia	81.77	77.68	84.76	*93.33*	91.12	85.03	83.59	73.52	83.81	*93.34*	90.65	86.93
MLPNN	83.69	83.91	83.81	93.33	*94.86*	83.02	82.89	68.93	87.14	90.66	*95.79*	84.21
FuzzyKNN + SVM	80.25	82.46	85.24	92.67	*96.73*	95.83	81.56	83.64	88.09	*96.7*	94.86	90.87

Table 3
Supervised discretization with FKNNSVM, $k = 10$

Discretization/dataset	Accuracy (%)	
	FIMDL	Kononenko
Ecoli	85.12	85.42
Glass	100	99.53
Vowel	87.25	87.25
Iris	94	94
Segmentation	87.7	90
Opt-digit	96.55	96.44
Average accuracy (%)	91.77	92.11

of numeric attributes of various types, which first converted to nominal attributes and then used to test the discretization algorithm. All the experiments are performed in an Intel core i5 processor with 2.3 GHz CPU, 1TB HDD, 4 GB RAM PC in Java Environment. The statistical significance is tested using Statistical software.

In order to evaluate the discretization carried out by the proposed algorithm with respect to other algorithms in the literature, We compared our methods: equal-width with fixed number of bins (we use 10 for our experiments), equal-frequency with fixed number of bins (we use 10 for the experiments), for class dependent (CD) and class independent (CI) scenario, as a part of unsupervised discretization and then Fayaad-Irani MDL and Kononenko as two supervised discretization scenarios, for building our granular computing model. The class will be unset temporarily before the granulation is applied in case of class independent and can be set for class dependent scenarios respectively. The classification is then obtained using hybrid fuzzy k-nearest neighbor support vector machine (FKNNSVM) with $k = 10$ in a 10-fold cross-validation methodology. The results on the test folds were compared through a Kruskal-Wallis one-way analysis of variance and Wilcoxon sign rank test as regards cross-validated predictive accuracy, as a test for statistical significance of the proposed model.

The results obtained with unsupervised and supervised information granulation process with various algorithms are presented in Tables 2–4. It can be observed that supervised Kononenko's discretization techniques produces more accurate classification accuracy since the partitions they produce are directly related to the class to be predicted and faster, taking less time for building the classification model. But, RRSE and the coverage is better in case of unsupervised EWD discretization using Fuzzy-KNN-SVM (FKNNSVM) for Class dependence (CD). Henceforth, we compare our equal width unsupervised (EWD) discretization and Kononenko's supervised discretization for information granulation along with various classifiers for their improved accuracy in comparison to their counterparts, which can be observed from Tables 2 and 3 respectively.

Table 4 provides us an insight about the performance of the various proposed methodology with all the used six datasets in terms of root relative square error, time taken to build the model, coverage and accuracy.

Table 5 presents the comparison on best accuracy for all dataset, with others. It is quite evident from Table 5 that our proposed methodologies with supervised and unsupervised (i.e. Kononenko and equal width) outperforms all existing works as mentioned. Further, we compare our approach with others granular computing approaches using same data sets that are highlighted in Table 6. It can be seen that, our proposed methodology is better in case of Ecoli, glass, iris and segmentation data and very close to opt-digit data and same with MLP or close with FKNNSVM with vowel dataset.

Simulations are carried out using data sets that do not have any missing attribute values and the obtained results observed from Tables 2–6; shows the effectiveness of our proposed Kononenko based granulation with FKNNSVM hybrid in comparison to the results obtained from many researchers [15,18,32,37,43,46].

For further research in this direction, post-operative patient data obtained from UCI machine repository with missing attribute values is used. An information/decision system is incomplete in case some values of conditional attributes from A are not known. Analysis of systems with missing values requires a decision on how to treat missing values. In plain words, objects with missing values are in a sense absorbed by close to them granules and missing values are replaced with most frequent values in objects collected in the granule. Then a comparison is made with the results obtained by the authors [3] using machine learning pro-

Table 4
Unsupervised CI and CD EWD and supervised discretization in terms of accuracy, RRSE and time

Algorithm/dataset			Vowel	Iris	Segmentation	Glass	Ecoli	Opt-digit	Average
FURIA	CI	ACC (%)	83.01	93.34	84.76	90.65	73.52	86.93	85.37
		RRSE (%)	59.88	41.69	65.22	40.29	78.48	49.73	55.88
		Coverage (%)	87.94	95.34	90.95	96.73	81.55	95.61	91.35
		Time (seconds)	0.57	0.03	0.13	0.02	0.18	8.73	1.61
	CD	ACC (%)	83.58	93.34	83.81	90.65	73.52	86.93	85.31
		RRSE (%)	58.45	41.69	52.62	40.29	78.48	49.73	53.54
		Coverage (%)	88.98	95.34	92.86	96.73	81.55	95.61	91.85
		Time (seconds)	0.71	0.01	0.09	0.02	0.18	9.29	1.72
MLPNN	CI	ACC (%)	87.71	90.66	87.14	95.79	72.63	84.21	86.36
		RRSE (%)	51.99	47.37	49.79	33.43	85.66	47.29	52.59
		Coverage (%)	94.14	94.67	93.34	97.19	95.24	90.21	94.13
		Time (seconds)	13.68	2.47	59.93	20.45	499.24	352.62	158.1
	CD	ACC (%)	82.89	90.66	87.14	95.79	68.93	84.21	84.94
		RRSE (%)	57.73	47.37	49.79	33.43	85.66	47.29	53.54
		Coverage (%)	93.68	94.67	93.34	97.11	95.24	90.21	94.04
		Time (seconds)	8.51	2.45	73.88	20.51	450.72	352.62	151.45
FKNNSVM, $K = 10$	CI	ACC (%)	86.68	96.7	85.72	95.12	83.04	91.2	89.74
		RRSE (%)	94.42	60.55	110.63	93.31	97.92	45.03	83.64
		Coverage (%)	100	100	99.53	100	98.81	100	99.72
		Time (seconds)	0.4	0.12	0.97	0.27	1.77	12.23	2.63
	CD	ACC (%)	81.56	96.7	88.09	94.86	83.64	90.87	89.29
		RRSE (%)	94.57	60.55	87.23	93.31	97.92	45.03	79.77
		Coverage (%)	100	100	100	100	98.81	100	99.80
		Time (seconds)	0.31	0.05	0.37	0.25	1.77	12.23	2.49
Kononenko + Furia		RRSE (%)	71.97	40.62	90.94	0	63.34	46.58	52.24
		Coverage (%)	84.67	94.67	73.87	100	87.20	95.32	89.29
		Time (seconds)	0.01	0.01	0.01	0.01	0.05	6.17	1.04
Kononenko + MLPNN		RRSE (%)	54.70	41.74	97.02	4.01	61.95	33.74	48.86
		Coverage (%)	96.16	98	65.24	100	88.9	98.05	91.06
		Time (seconds)	65.25	0.38	122.38	2.61	673.58	846.8	285.17
Kononenko + FKNNSVM, $k = 10$		RRSE (%)	85.23	62.72	90.94	92.71	97.81	90.73	86.69
		Coverage (%)	99.71	100	68.73	100	98.81	100	94.54
		Time (seconds)	0.98	0.09	2.75	0.52	1.78	3.99	1.68

Table 5
Comparison on best accuracy for all dataset, with others

Discretization	Algorithm	Average accuracy (%)	Discretization	Algorithm	Average accuracy (%)	Discretization	Algorithm	Average accuracy (%)
EFD	FURIA (ours)	85.61	EWD	FURIA (ours)	85.31	konenko	FURIA (ours)	82.31
	MLPNN (ours)	87.1		MLPNN (ours)	84.90		MLPNN (ours)	85.02
	FNNSVM (ours)	*88.86*		FNNSVM (ours)	*89.28*		FNNSVM (ours)	*91.77*
	C4.5 [43]	73.04		C4.5 [43]	72.52	FI-MDL	C4.5 + MDL [43]	74.44
	KNN [43]	75.57		KNN [43]	46.80		KNN + MDL [43]	70.02
	NB [43]	74.91		NB [43]	*75.7*		NB + MDL [43]	73.69
	GSC +MD + I2BG [37]	87.9				Rough mereology [32]		88.19

gram LERS_LB 2.5 in knowledge acquisition for expert system development in nursing. While the proposed hybrid Kononenko + FKNNSVM achieves an accuracy of 67.78%, the latter has only 48%. It is also observed that while using Kononenko + Fuzzy KNN without SVM, the accuracy is reduced to 65.52%, from 67.78%.

Finally, in order to measure the suitability of classifiers over one another in terms of accuracy, nonparametric Kruskal-Wallis and Wilcoxon sign rank test are performed with a risk level $\alpha = 0.05$. The result obtained are shown in Tables 7 and 8.

(i) The *Kruskal-Wallis one-way analysis of variance* by ranks (named after William Kruskal

Table 6
Comparison with others granular computing approaches

Dataset/Algorithm	Ecoli	Glass	Iris	Vowel	Segmentation	Opt_digit
FRGNN [46]	–	–	–	*87.71*	–	–
RFMLP [46]	–	–	–	86.55	–	–
FRGNN [46]	–	–	–	86.55	–	–
DGrC [18]	–	80.1	–	–	–	–
Neuro-Fuzzy GrC [15]	–	–	–	–	–	*98.16*
FKNNSVM + CI (ours), $k = 10$	83.04	94.86	*96.7*	86.68	85.72	90.87
Kononenko + FKNN SVM (ours)	*85.42*	*99.53*	94	87.25	*90*	96.44
MLPNN + CI (ours)	72.63	95.79	90.66	*87.71*	87.14	84.21

Table 7
Statistical significance test on accuracy (Kruskal-Wallis test H takes ties into account)

H (observed value)	3.644
H (critical value)	11.070
DF	5
One-tailed p-value	0.602
Alpha	0.05

The Kruskal-Wallis H is distributed as a Chi-square.
At the level of significance Alpha = 0.050 the decision is to not reject the null hypothesis of absence of difference between the 6 samples. In other words, the difference between the samples is not significant.
From all comparisons, it is quite evident that our granular computing approach using FKNNSVM with EWD is equally significant compared to the other methods.

Table 8
Statistical significance test on accuracy (Wilcoxon signed-ranks test/two-tailed test)

T	5.000
T (expected value)	7.500
T (variance)	13.750
Z (observed value)	−0.674
Z (critical value)	1.960
Two-tailed p-value	0.500
Alpha	0.05

The normalized Wilcoxon's T is tested against the normal distribution.
At the level of significance Alpha = 0.05, the decision is not to reject the null hypothesis that the samples are not different. In other words, the difference between the samples is not significant.
This shows that the proposed Kononenko FKNN with SVM hybrid is also equally significant.

and W. Allen Wallis) is a non-parametric method for testing whether samples originate from the same distribution [23]. It is used for comparing two or more samples that are independent, and that may have different sample sizes, and extends the Mann-Whitney U test to more than two groups. The parametric equivalent of the Kruskal-Wallis test is the one-way analysis of variance (ANOVA). When rejecting the null hypothesis of the Kruskal-Wallis test, by that time, at least one of sample stochastically dominates at least one other sample.

(ii) *The Wilcoxon Signed Rank* procedure assumes that the sample under investigation is randomly taken from a population, with a symmetric frequency distribution. The symmetric assumption does not assume normality, simply that there seems to be roughly the same number of values above and below the median. The Wilcoxon procedure computes a test that is compared to an expected value, which is computed by summing the ranked differences of the deviation of each variable from a hypothesized median above the hypothesized value.

6. Conclusions and future work

In this paper, an attempt was made to show how the discretization based granular Fuzzy-KNN-SVM can design an effective classification model with high accuracy. Our experimental results indicate that with 7 public datasets, on an average, Fuzzy-KNN-SVM with EWD is better than other unsupervised granulation algorithm and for supervised granulation, Kononenko finds a better place in comparison to Fayyad and Irani's Minimum Description Length (MDL). Since most of the state of the art classifiers are performing well on these datasets, it is clear that the data transformation is more important than the classifier itself.

The experimental results with the proposed methodology on the average gives better performance in comparison to others. Finally, we validate our method for the development of parsimonious models using non-parametric statistical test that could be set as a benchmark for statistical classifiers.

Even though our proposed supervised and unsupervised granulation with Hybrid Fuzzy KNN and SVM has obvious advantages compared to traditional SVM and many other proposed approach available in the recent literature, but the effective application in the daily life with better data granulation still is an important aspect of future research.

References

[1] A. Bargiela and W. Pedrycz, The roots of granular computing, *Proc of the IEEE Int'l Conf on Granular Computing* (2006), 806–809.

[2] A. Bargiela and W. Pedrycyz, Towards a theory of granular computing for human-centered information processing, *IEEE Transaction on Fuzzy Systems* **16**(2) (2008), 320–330.

[3] A. Budihardjo, J. Grzymala-Busse and L. Woolery, Program LERS_LB 2.5 as a tool for knowledge acquisition in nursing, *Proceedings of the 4th Int, Conference on Industrial and Engineering Applications of AI and Expert Systems* (1991), 735–740.

[4] A. Ganivada, S. Dutta and S.K. Pal, Fuzzy rough granular neural networks, *Fuzzy Granules and Classification, Theoretical Computer Science*, Elsevier **412** (2011), 5834–5853.

[5] A. Ganivada, S.S. Ray and S.K. Pal, Fuzzy rough set and a granular neural network for unsupervised feature selection, *Neural Networks*, Elsevier **48** (2013), 91–108.

[6] A. Kazem, E. Sharifi, F.K. Hussain, M. Saberi and O.K. Hussain, Support vector regression with chaos-based firefly algorithm for stock market price forecasting, *Applied Soft Computing*, Elsevier **13** (2013), 947–958.

[7] A. Skowron and J. Stepaniuk, Information granules: Towards foundations of granular computing, *International Journal of Intelligent Systems* **16** (2001), 57–85.

[8] C.H. Wan, L.H. Lee, R. Rajkumar and D. Isa, A hybrid text classification approach with low dependency on parameter by integrating K-nearest neighbour and support vector machine, *Expert Systems with Applications* **39** (2012), 11880–11888.

[9] C. Lin, C. Yeh, S. Liang, J. Chung and N. Kumar, Support-vector-based fuzzy neural network for pattern classification, *IEEE Transactions on Fuzzy Systems* **14**(1) (2006), 31–41.

[10] D. Li, D. Miao and H. Zhang, The theory, model and method of granular computing, *Journal of Fudan (Natural Science)* **43**(5) (2004), 837–841.

[11] D. Li, D. Miao, D. Zhang et al., An overview of granular computing, *Computer Science* **32**(9) (2005), 1–12.

[12] D. Li, Z. Chen and J. Yin, A new attribute reduction recursive algorithm based on granular computing, *Journal of Computers* **8**(3) (Mar 2013), 630–637.

[13] S. Ding, L. Xu, H. Zhu and L. Zhang, Research and progress of cluster algorithms based on granular computing, *International Journal of Digital Content Technology and Its Applications* **4**(5) (Aug 2010), 96–104.

[14] E.A. Zanaty, Support vector machines (SVMs) versus multilayer perception (MLP) in data classification, *Egyptian Informatics Journal* **13** (2012), 177–183.

[15] G. Panoutsos and M. Mahfouf, Discovering knowledge and modelling systems using granular computing and Neuro-fuzzy structures, in: *NiSIS'05, European Symposium on Nature-Inspired Smart Information Systems*, Albufeira, Portugal, (2005).

[16] H. Huang, S. Ding, F. Jin, J. Yu and Y. Han, A novel granular support vector machine based on mixed kernel function, *International Journal of Digital Content Technology and its Applications (JDCTA)* **6**(20) (2012), 484–492. doi: 104156/jdcta.vol6.issue20.52.

[17] H. Li and S. Ding, Research and development of granular neural networks, *Applied Mathematics and Information Sciences Journal* **7**(3) (2013), 1251–1261.

[18] H. Liu, C. Liu and C.-A. Wu, Granular computing classification algorithms based on distance measures between granules from the view of set, *Computational Intelligence and Neuro Science*, Hindawii Publishing Corporation, (2014), ID-656790, 1–9,

[19] I. Kononenko and H.S. June, Attribute selection for modelling, *Future Generation Computer Systems* **13**(2–3) (1997), 181–195.

[20] I. Mitov, Kr. Ivanova, Kr. Markov, V. Velychko, K. Vanhoof and P. Stanchev, PaGaNe – A classification machine learning system based on the multidimensional numbered information spaces, in: *Proc of Intelligent Systems and Knowledge Engineering*, Hasselt, Belgium **11**(27–28) (2009), 279–287.

[21] I.U. Sikder, A granular computing approach to decision analysis using Rough set theory, in: *Proc of the 2010 International Conference on Industrial Energy and Operations Management*, Dhaka, Bangladesh, (Jan 9–10 2010). http://citeseerx.ist.psu.edu/viewdoc/summary?doi=101.1.397.6626.

[22] J.C. Huehn and E. Huellermeier, FURIA: An algorithm for unordered fuzzy rule induction, *Data Mining and Knowledge Discovery* **19**(3) (2009), 293–319.

[23] J.D. Spurrier, On the null distribution of the kruskal – wallis statistic, *Journal of Nonparametric Statistics* **15**(6) (2003), 685–691.

[24] J. Dougherty, R. Kohavi and M. Sahami, Supervised and unsupervised discretization of continuous features, *Proceedings of Twelfth International Conference on Machine Learning* **95**(10) (1995), 194–202.

[25] J. Hu and C. Guan, An emotional agent model based on granular computing, *Mathematical Problems in Engineering*, Hindawii Publishing Corporation, (2012), Article ID 601295, doi:101155/2012/601295..

[26] J.T. Yao and Y.Y. Yao, A granular computing approach to Machine Learning, in: *Proc of the 1st International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'02)* (2002), 732–736.

[27] J.T. Yao, A ten year review of granular computing, in: *Proc of IEEE International Conference on Granular Computing*, silicon Valley, USA, (Nov 2–5 2007), 734–739.

[28] L.A. Zadeh, Some reflections on information granulation and its centrality in granular computing, computing with words, the computational theory of perceptions and precisiated natural language, in: *Proc Data Mining, Rough Sets and Granular Computing, Studies in Fuzziness and Soft Computing* **95** (2002), 3–20.

[29] L.A. Zadeh, Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems* **19** (1997), 111–127.

[30] L.A. Zadeh, Fuzzy sets and information granularity, in: *Advances in Fuzzy Set Theory and Applications* N. Gupta, R. Ragade and R. Yager, eds, North-Holland, Amsterdam, 1979, pp. 3–18.

[31] L.A. Zadeh, Some reflections on soft computing, granular computing and their roles in the conception, design and utilization on information, *Intelligent Systems, Soft Computing*, Springer-Verlag **2** (1998), 23–25.

[32] M.A. Mahmood, N. El-Bendary, A.E. Hassanien and H.A. Hefny, *Classification Approach based on Rough Mereology, ISI-2013*, India, 265–273.

[33] M. Zhang and J.-X. Cheng, Pattern classification with granular computing, *GESTS International Transaction on Computer Science and Engineering* **20**(1) (2005), 137–145.

[34] M.G.C.A. Cimino, B. Lazzerini, F. Marcelloni and W. Pedrycyz, Genetic interval neural networks for granular data regression, *Information Sciences*, Elsevier **257**(9) (2014), 313–330.

[35] M.K. Ismail and V. Ciesielsk, An empirical investigation of

the impact of discretization on common data distributions, design and application of hybrid intelligent systems, in: *Proc of Third International Conference on Hybrid Intelligent Systems* **105** (2003), 692–701.
[36] N.B.H. Ahmad, S.M. Shamsuddin and A. Abraham, Granular mining of student's learning behavior in learning management system using rough set technique, *Computational Intelligence for Tech Enhanced Learning*, F. Xhafa et al., eds, Springer-Verlag Berlin Heidelberg, (SCI 273) (2010), 99–124.
[37] P. Zhu and Q. Hu, Adaptive neighbourhood granularity selection and combination based on margin distribution optimization, *Information Aciences*, Elsevier **249** (2013), 1–12,
[38] R. Abraham, J.B. Simha and S.S. Iyengar, Effective discretization and hybrid feature selection using Naïve bayesian classifier for Medical data mining, *International Journal of Computational Intelligence Research* **5**(2) (2009), 116–129.
[39] R. Singh, M. Vatsa and A. Noore, Multiclass m_Granular soft support vector machine: A case study, in: *Proceedings of International Conference on Pattern Recognition* (2008), 1–4.
[40] S.B. Devasenapati and K.I. Ramachandran, Random forest based misfire detection using kononenko discretiser, *ICTACT Journal on Soft Computing* **2**(2) (2012), 270–275.
[41] S. Ding and B. Qi, Research of granular support vector machine, *Artif Intell Rev* **38**(1–7) (2012), Springer, DOI: 101007/s10462-011-9235-9.
[42] S. Fong, O. Mohammed, J. Fiaidhi, S. Mohammed and C.K. Kwoh, Measuring similarity by prediction class between biomedical datasets via fuzzy unordered rule induction, *International Journal of Bio-Science and Bio-Technology* **6**(2) (2014), 159–168. http://dxdoi.org/10.14257/ijbsbt.2014.6.2.16.
[43] S. Garcıa, J. Luengo, J.A. Saez, V. Lopez and F. Herrera, A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning, *IEEE Transactions on Knowledge and Data Engineering* **25**(4) (2013), 734–750.
[44] S. Kotsiantis, D. Kanellopoulos and D. Techniques, A recent survey, *GESTS International Transactions on Computer Science and Engineering* **32**(1) (2006), 47–58.
[45] S.K. Pal, Granular mining and rough-fuzzy pattern recognition: A way to natural computation, *IEEE Intelligent Informatics Bulletin* **13**(1) (2012), 3–13.
[46] S.K. Pal and S.K. Meher, Natural computing: A problem solving paradigm with granular information processing, *Applied Soft Computing*, Elsevier **13** (2013), 3944–3955.
[47] S.K. Pal, B.U. Shankar and P. Mitra, Granular computing, rough entropy and object extraction, pattern recognition letters, Elsevier **26** (2005), 2509–2517.
[48] T.Y. Lin, Data mining and machine oriented modelling: A granular computing approach, *Journal of Applied Intelligence*, Springer **13**(2) (Oct 2000), 113–124.
[49] U.M. Fayyad and K.B. Irani, Multi-interval discretization of continuous valued attributes for classification learning, *Proceedings of the International Joint Conference on Uncertainty in Artificial Intelligence* **2**(1) (1993), 1022–1027.
[50] W. Pedrycz, Granular computing, *An Emerging Paradigm*, ed, Physica-Verlag, Heidelberg, ISSN, (2001), 1434–9922.
[51] W. Pedrycyz, Granular computing and logic processing in intelligent systems, *AI-METH 2003-AI methods* (Nov 5–7 2003), Gliwice, Poland, http://www.academia.edu/3056512/Granular_Computing_and_Logic_Processing_in_Intelligent_Systems.
[52] W. Pedrycyz, Associations among information granules and their optimization in granulation-degranulation mechanism of granular computing, *International Journal of Fuzzy Logic and Intelligent Systems* **13**(4) (Dec 2013), 245–253.
[53] W. Wang, H. Guo, J.Y. Feng and J. Bi, Granular support vector machine based on mixed measure, *Neuro Computing* **101** (2013), 116–128.
[54] Y. Chen, D. Miao and R. Wang, Outlier detection based on granular computing, in: *RSCTC 2008*, C.-C. Chan et al., eds, (LNAI 5306), Springer-Verlag, berlin, Heidelberg, (2008), 283–292.
[55] Y. Tang, Y.-Q. Zhang, Z. Huang, X. Hu and Y. Zhao, Recursive fuzzy granulation for gene subsets extraction and cancer classification, *IEEE Transaction on Information Technology in Biomedicine* **12**(6) (Nov 2008), 723–730.
[56] Y. Wang, On system algebra: A denotational mathematical structure for abstract systems modelling, *International Journal of Cognitive Informatics and Natural Intelligence*, IGI Publishing, USA **2**(2) (Apr 2008), 20–43.
[57] Y. Wang, L.A. Zadeh and Y. Yao, On the system algebra foundations for granular computing, *International Journal of Software Science and Computational Intelligence* **1**(1) (Jan–Mar 2009), 64–86.
[58] Y.Y. Yao, A partition model of granular computing, *Transactions on Rough Sets*, Springer **1** (LNCS 3135) (2004), 232–253.
[59] Y.Y. Yao, Granular computing: Basic issues and possible solutions, *Proceedings of the 5th Joint Conference on Information Sciences* (2000), 186–189.
[60] Y.Y. Yao and N. Zhong, Granular computing using information tables, *Data Mining, Rough Sets and Granular Computing Studies in Fuzziness and Soft Computing* **95** (2002), 102–124.
[61] Y. Yang and I.G. Webb, A comparative study of discretization methods for naïve bayes classifiers, in: *Proceedings of PKAW*, Japan, (2002), 159–173.
[62] Z.I. Mooe and A. Inoue, Effectiveness of value granulation in machine learning for massively large and complex domain, in: *Proc of IPMU 2008*, L. Magdalena et al., eds, 22–27 Jun 2008, pp. 855–862.
[63] Z. Pawlak, Granularity of knowledge, indiscernibility and rough sets, *Proceedings of IEEE International Conference on Fuzzy Systems*, (1998), 106–110.