Multigranulation Supertrust Model for Attribute Reduction

Weiping Ding, Senior Member, IEEE, Witold Pedrycz, Fellow, IEEE, Isaac Triguero, Member, IEEE, Zehong Cao, Member, IEEE, and Chin-Teng Lin, Fellow, IEEE

Abstract—As big data often contains a significant amount of uncertain, unstructured, and imprecise data that are structurally complex and incomplete, traditional attribute reduction methods are less effective when applied to large-scale incomplete information systems to extract knowledge. Multigranular computing provides a powerful tool for use in big data analysis conducted at different levels of information granularity. In this article, we present a novel multigranulation supertrust fuzzy–rough set-based attribute reduction (MSFAR) algorithm to support the formation of hierarchies of information granules of higher types and higher orders, which addresses newly emerging data mining problems in big data analysis. First, a multigranulation supertrust model based on the valued tolerance relation is constructed to identify the fuzzy similarity of the changing knowledge granularity with multimodality attributes. Second, an ensemble consensus compensatory scheme was adopted to calculate the multigranular trust degree based on the reputation at different granularities to create reasonable subproblems with different granulation levels. Third, an equilibrium method of multigranular coevolution is employed to ensure a wide range of balancing of exploration and exploitation, and this strategy can classify super elitists’ preferences and detect noncooperative behaviors with a global convergence ability and high search accuracy. The experimental results demonstrate that the MSFAR algorithm achieves a high performance in addressing uncertain and fuzzy attribute reduction problems with a large number of multigranularity variables.

Index Terms—Ensemble consensus compensatory scheme, fuzzy–rough attribute reduction, multigranulation supertrust model, valued tolerance relation.

I. INTRODUCTION

RECENTLY, there has been explosive growth in the amount of data generated, and the term “big data” is being used to refer to the challenges of handling data with a high volume, variety, velocity, intrinsic value, and uncertain veracity. These “five V’s” are the key features defining the essence of big data [1], [2]. Big data has attracted considerable attention from a variety of circles of scientific research, marketing, business management, and government decision making, leading to an upsurge of research [3]–[8]. Although a large candidate set of attributes is provided in big data problems, most may be redundant or irrelevant, which highly diminishes the learning performance of decision-making algorithms. The complexity of the big data problem mainly arises from the very large number of decision variables and various types of constraints. Thus, it has become highly desirable to develop some effective attribute reduction (feature selection) methods to extract useful knowledge hidden in large-scale data repositories. Since big data can often be incomplete, uncertain, and vague in reality, conventional knowledge discovery techniques, ranging from models, algorithms, and systems to applications, have been challenged in terms of how to store, manage, process, and analyze the complex attribute sets of big data [9]–[11].

With the increase in big data, researchers have started realizing the existence of data space alongside natural and social spaces and shown remarkable interest in its exploration. Structuralized knowledge organization and reasoning is considered an effective paradigm for handling large-scale tasks. In the recent past, a considerable amount of work has focused on a new research area—granular computing (GrC)—which is situated against the background of other human-centered information processing paradigms. GrC, which is a term coined by Zadeh [12], refers to a new knowledge representation and reasoning paradigm with information granules. Fuzzy sets and rough sets are the two main formal frameworks among active branches of GrC [13], which are of vital importance for the understanding of big data analysis completed at different granularity levels. They provide two powerful conceptual and algorithmic vehicles for multiple-view data analysis. The fuzzy set theory introduced by Zadeh [14] in 1965 is a formal mechanism by which to
represent and manipulate concepts with ambiguous boundaries and to understand and apply the processes employed in human reasoning. However, a fuzzy set is characterized by only a membership function, which ignores uncertain information and thus degrades its performance in big data analysis. The rough set theory proposed by Pawlak [15] in 1982 has been applied to contend with uncertainty caused by indiscernibility and incompleteness [16]–[21]. Since the rough set theory is complementary to the fuzzy set theory, fuzzy–rough sets have appeared as a newly emerging combination delivering the advantages of both complementary areas and are considered to provide a more powerful model for analyzing uncertainty in big data [22]–[24]. A fuzzy–rough set is defined by two fuzzy sets, fuzzy upper and lower approximations, which are obtained by extending the corresponding notions of a rough set.

In the fuzzy–rough framework, elements have membership grades located within some range, which allows for greater flexibility in handling uncertain information [25]. In the Boolean case, elements that belong to the lower approximation are represented as belonging to the approximated set with absolute certainty. Therefore, it has become timely and strongly justified to develop effective fuzzy–rough set algorithms with multigranulation to enhance understanding and reasoning in big data analytics. Type-2 fuzzy sets, as a higher type with a higher order of information granules, extend the expressive capabilities of Type-1 fuzzy sets, and they are able to represent the imprecision of the membership function of fuzzy sets. A Type-2 interval number (IN) is a mathematical object that can be interpreted either probabilistically or possibilistically. The use of an IN is particularly appropriate when modeling linguistic concepts [26], [27]. Type-2 fuzzy sets have the potential to model uncertainties despite the large number of associated computations, especially when applied to nonreal-time applications [28], [29]. However, Type-2 fuzzy sets do not place any constraints on the continuity or other properties of their embedded sets.

Fuzzy–rough set models provide a method by which discrete or a real-valued noisy data or a mixture of both can be greatly reduced so that it can be effectively applied to both regression and classification of large-scale datasets. Fuzzy–rough set research has attracted considerable attention in recent years. Some approaches have been proposed to improve the performance of traditional fuzzy sets and rough sets as follows. Wang et al. [30] presented a new nearest-neighbor clustering classification algorithm based on the fuzzy–rough set theory, in which every training sample was made according to fuzzy roughness, and then training sample points in a class boundary or overlapping regions were removed. Hassanian [31] introduced a hybrid scheme in conjunction with statistical feature extraction techniques by combining the advantages of both rough sets and fuzzy sets, wherein rough sets were employed for the reduction generation of the minimal number of features, and fuzzy sets were considered as an image preprocessing technique to enhance the contrast of the whole image. It was reported, however, that fuzzy–rough sets are sensitive to noisy samples. To alleviate this shortcoming, Hu et al. [32] discussed why the models of rough sets are sensitive to noise and developed some robust fuzzy–rough set models based on fuzzy lower approximations. Petrosino and Salvi [33] presented a multiscale algorithm based on rough fuzzy sets in which rough sets handled the vagueness and fuzzy sets handled the coarseness. Sarkar [34] generalized the concept of rough membership functions to rough–fuzzy membership functions, wherein the value signified the rough uncertainty as well as the fuzzy uncertainty associated with the pattern. An et al. [35] proposed a novel robust data-distribution-aware fuzzy–rough set model by computing lower and upper approximations. However, the proposed models cannot be used to handle multimodal big data in real-world applications. Xu et al. [36] put forward a novel data redundancy reduction approach based on both the fuzzy–rough set theory and the information theory. Salama [37] provided fuzzy–rough attribute reduction software to facilitate the reduction of high-dimensional data. Zeng et al. [38] proposed the fuzzy–rough set approach for incremental feature selection in hybrid information systems. Maji and Garai [39] presented an interval type-2 (IT2) fuzzy–rough feature selection method, judiciously integrating the merits of the IT2 fuzzy set and rough sets to effectively reduce real-valued noisy features. Zhao et al. [40] analyzed the nested topological structure of fuzzy–rough sets with incremental parameters and designed a novel algorithm to compute a nested classifier by reflecting all possible parameters. Peng and Mi [41] considered the multigranulation fuzzy–rough sets of an information system by the minimal and maximal membership degrees based on multifuzziness tolerance relations. Yang et al. [42] presented two incremental algorithms for attribute reduction with fuzzy–rough sets for one and multiple incoming samples. Wang et al. [43] introduced a fitting fuzzy–rough set model to guarantee the maximal membership degree of a sample to its own category. Because fuzziness is employed in the rough set theory, more reduction information relevant to continuous attributes can be successfully acquired. Hu et al. [50] proposed a model of multikernel fuzzy–rough sets and described a parallel strategy to handle large-scale multimodality fuzzy data attribute reduction.

Although fuzzy–rough attribute reduction methods have shown promising performance, they cannot cope well with the multimodality of big data and a large variety of real-world applications that involve the challenging complexity of big data. In practice, most attribute reduction and classification tasks are associated with mixtures of numerical and categorical attribute features. The size of multimodal big data is usually very large, resulting in extensive time consumption and the use of massive parallel processing databases in performing attribute reduction. Obviously, this can be greatly detrimental to the traditional attribute reduction performance for analyses of incomplete large-scale information systems. Furthermore, noisy attributes are also one of the main sources of uncertainty in big data applications. Although a few fuzzy–rough attribute reduction methods are robust toward complex noisy attributes, they require more user-supplied information, and there is a lack of continuity and inheritance in their internal relationships, which results in unsatisfactory performances. Meanwhile, when the volume of massive data objects increases in the database, much more computing time and space overhead are necessary to address the new rendering decision attributes. Currently, few works have considered the multigranulation algorithm for big
In the rough set theory, the universe is divided into a set of equivalence classes according to the attribute values of objects. An information system can be defined as a decision table by \( T = (U, C, D, V, f) \), where \( C \) is the set of condition attributes, \( D \) is the set of decision attributes, \( V \) is the value set of all attributes, and \( f : U \times \{ C \cup D \} \rightarrow V \) is the information function, such that \( f(x, a) \in V_a \) for each \( a \in \{ C \cup D \}, x \in U \).

**Definition 2** [16], [17]: For \( P \subseteq \{ C \cup D \} \), an equivalence relation \( IND(P) \) is defined as follows:

\[
IND(P) = \{(x, y) \in U \times U | \forall a \in P, f(x, a) = f(y, a)\}.
\] (1)

**Definition 3** [14]: Let \( \{ D_1, D_2, \ldots, D_r \} \) be a family of fuzzy sets on \( U \) if \( \sum_{i=1}^{r} D_i(x) = 1 (\forall x \in U) \), where \( \{ D_1, D_2, \ldots, D_r \} \) is considered a fuzzy partition. For \( \forall x \in U \), the fuzzy decision of \( x \) is defined by

\[
\hat{D}_i(x) = \frac{[x]_B \cap D_i}{[x]_B}, \quad i = 1, 2, \ldots, r, \quad \text{and } x \in U
\] (2)

where \( \hat{D}_i(x) \) denotes the fuzzy decision of \( x \) to \( D_i \).

**Definition 4**: For an object \( x \in U \), the fuzzy positive region of \( D \) relative to \( B \) is defined as

\[
POS_B(D)(x) = \bigcup_{i=1}^{r} B \hat{D}_i(x)
\] (3)

where \( \hat{D}_i(x) \) is a set of decision attributes.

**Definition 5** [42]: For each condition attribute \( a \in A \), one can define a fuzzy binary relation \( R_a \), which is called a fuzzy equivalence relation if \( R_a \) is reflexive \((R(x, x) = 1)\), symmetric \((R(x, y) = R(y, x))\), and sup-min transitive \((R(x, y) \geq \sup_{x \in U} \min (R(x, z), R(z, y)))\) for \( \forall x, y, z \in U \). A subset \( B \subseteq A \) can also define a fuzzy equivalence relation, denoted by \( R_B = \cap_{a \in B} R_a \). Based on the fuzzy equivalence relation, the concept of fuzzy rough set is defined as follows.

Let \( F(U) \) be the fuzzy power set of \( U \) and \( B \subseteq A \). For each \( x \in U \), a pair of lower and upper approximation operators of \( X \subseteq F(U) \) based on \( R_B \) is defined by

\[
\overline{R_B}(X)(x) = \inf_{u \in U} \max \{ 1 - R_B(x, u), X(u) \},
\] (4)

\[
\underline{R_B}(X)(x) = \sup_{u \in U} \min \{ R_B(x, u), X(u) \}
\] (5)

where the lower approximation \( \overline{R_B}(X)(x) \) is considered as the degree of \( x \) certainly belonging to \( X \), whereas the upper approximation \( \underline{R_B}(X)(x) \) is the degree of \( x \) possibly belonging to \( X \).

\((\overline{R_B}(X), \underline{R_B}(X))\) is referred to as the fuzzy rough set of \( X \) with respect to \( B \), which is defined based on Max \( r \)-conorm and Min \( r \)-norm.

**Definition 6**: For the fuzzy–rough attribute reduction process, it is necessary to determine the dependence degree of the decision features. The dependence function of \( D \) relative to \( B \)
is formally described by
\[
\partial_B(D) = \frac{\sum_{x \in U} \text{POS}_B(D)(x)}{|U|}
\]  
(6)
where \(0 \leq \partial_B(D) \leq 1\), and it is defined as the ratio of the sizes of the positive region relative to all samples in the feature space.

**Definition 7:** The main thought behind the attribute reduction process using fuzzy–rough sets is to find a minimal subset of attributes that keeps the positive region unchanged, so that those wiped features will not affect the decision making. Considering a decision table \((U, A \cup D)\), a subset \(\text{Red} \subseteq A\) is called a reduct of \(A\) relative to \(D\) if the following conditions are satisfied

1. for \(\forall x \in U\), \(\text{Pos}_A(D)(x) = \text{Pos}_\text{Red}(D)(x)\);
2. \(\forall a \in P\), \(\exists y \in U\), \(\text{Pos}_{P \setminus \{a\}}(D)(y) = \text{Pos}_\text{Red}(D)(y)\).
(7)
(8)

Notice that the reduct is usually not unique. Let \(\text{Red}_D(A)\) denote the set of all reducts with respect to \((U, A \cup D)\), and then \(\text{Core}_D(A) = \cap \text{Red}_D(A)\) is called the core of \((U, A \cup D)\). It is easier to obtain the core first and then find a reduct based on the core.

**Definition 8 [44, 45]:** In the improved quantitative tolerance model, for an incomplete information system \(\text{DIIS} = \{U, AT, V, f\}\), \(a \in B \subseteq AT\), \(V^a = \{(k^a_1, v_1), (k^a_2, v_2), \ldots, (k^a_m, v_m)\}\), where \(v_1, v_2, \ldots, v_m\) are all the possible values of the attribute \(a\). \(k^a_i = |\{x \in U|a(x) = v_i\}|\), where \(\bullet\) denotes the cardinality of a set. \(\forall x \in U, a(x) = v_i\), and its probability is \(k^a_i/(k^a_1 + k^a_2 + \cdots + k^a_m)\). \(\forall x, y\), their similarity degree based on the value of \(a \in A\) is defined as

\[
\text{VS}_{[a]}(x, y) =
\begin{cases}
1, & a(x) = a(y) \land a(x) \neq \bullet \land a(y) \neq \bullet \\
\frac{k^a_i}{\left(\sum_{j=1}^m k^a_j\right)}, & a(x) = v_i \land a(y) \\
0, & a(x) \neq a(y) \land a(x) \neq \bullet \land a(y) \neq \bullet \\
\frac{\sum_{j=1}^m k^a_i}{\left(\sum_{j=1}^m k^a_j\right)}, & a(x) = \bullet \land a(x) = \bullet \\
\frac{\sum_{i=1}^m k^a_i}{\left(\sum_{i=1}^m k^a_i\right)}, & a(x) = \bullet \land a(x) = \bullet \\
\end{cases}
\]  
(9)

Therefore, the similarity between \(x\) and \(y\) in \(A \subseteq AT\) is defined as
\[
\text{VT}_A(x, y) = \prod_{v_i \in A} \text{VT}_{v_i}(x, y)
\]  
(10)
and the quantitative tolerance class of \(x\) in \(A \subseteq AT\) is denoted by
\[
\text{VT}^{w}_{A}(x) = \{y \in U : \text{VT}_A(x, y) \geq w\} \cup \{x\}.
\]  
(11)

**Definition 9:** A multigranular valued tolerance relation is considered a good decomposition fuzzy–rough set framework for addressing large-scale problems with dynamically increasing complexity. For the information system \(\text{DIS} = \{U, A, V, f\}\), \(A_1, A_2, \ldots, A_m \subseteq A\), which corresponds to the classification threshold with a sequence of \(m\) sequence of attribute sets. \(\forall X \subseteq U\), where the upper approximation and lower approximation based on the grade multigranulations are defined as follows:

\[
\sum_{i=1}^m A^0_i w_i X = \{x \in U : V^{A_i} \subseteq X \lor V^{A_2} \subseteq X \lor \ldots \lor V^{A_m} \subseteq X\},
\]  
(12)
\[
\sum_{i=1}^m A^0_i w_i X = \sim \sum_{i=1}^m A^0_i w_i (\sim X).
\]  
(13)

Therefore, \((\sum_{i=1}^m A^0_i w_i X, \sum_{i=1}^m A^0_i w_i X)\) is represented as the optimistic multigranularity fuzzy–rough set model based on quantitative tolerance.

### III. MULTIGRANULATION SUPERTRUST MODEL WITH A SELF-EVOLVING COMPENSATORY SCHEME

Traditional attribute reduction methods are satisfactory to a certain extent, but they are not capable of addressing massive amounts of complex large-scale data. Thus, there is a need for devising an effective trust model to efficiently handle the inherent multimodality attributes characteristics of big data. A multigranulation supertrust fuzzy–rough set model based on the valued tolerance relation is constructed to extract the fuzzy similarity of changing knowledge granularity for fuzzy classification. This model effectively solves the problem of missing data in an incomplete large-scale information system. With the increasing dimensionality of multigranulation space, most approaches to extract the fuzzy similarity of knowledge granularity are easily trapped in local optima due to overexploitation; therefore, their performance deteriorates. To achieve a better balance between the exploration and exploitation of knowledge granularity for solving complex large-scale datasets, we propose a novel multigranulation supertrust model with a self-evolving compensatory scheme to calculate the multigranular trust degree according to (17) at different granularities, thereby splitting the large data set into reasonable subdatasets. In addition, this model can explore the search space and locate the global best region during the fuzzy–rough attribute reduction process, as well as accelerate the premature convergence speed.

#### A. Multigranulation Supertrust Model

Since there are various methods for calculating the credibility of a population, a practical method aims to process the multigranulation supertrust framework to adjust trust relationships based on subpopulations' interactions in different granularity spaces. We construct a granu-subpopulation architecture according to the trust degree of evolutionary elitists, including super elitists, denoted by “•”, and ordinary elitists, denoted by “”. The credibility of subpopulations in the same granu-subpopulation is calculated according to the trust calculation mechanism based on their respective reputations, which has been proven to be a good reflection of trust relationships between subpopulations of different granularities. Since the interaction between elitists in the same granu-subpopulation will be more frequent than those between granu-subpopulations in the
common topology, the trust degree can be quickly established, and the granu-subpopulation can be evaluated effectively. The dynamic trust execution process is described in Fig. 1. Two types of supertrust relationships are employed to play two roles within different granu-subpopulations for fuzzy–rough attribute reduction. As specified in Fig. 2, the supertrust relationships between the super elitist and the ordinary elitist and between ordinary elitists are both direct trust relationships, and those between the super elitists within different granu-subpopulations are recommendation trust relationships. The main steps of the process are described in Algorithm 1.

**B. Ensemble Consensus Compensatory Scheme**

In this section, an ensemble consensus compensatory scheme is presented to evaluate the trust information of recommended granu-subpopulations. Its main novelty lies in calculating the multigranular trust degrees in different granularities based on the fuzzy granularity compensatory scheme, as well as addressing reasonable subproblem splitting. By analyzing the change in the knowledge granularity produced by coarsening and refining in the process of attribute reduction, the overall performance

**Algorithm 1: Multigranulation Super-Trust Model (MSTM):**

1. Set the number of granularity subpopulations $N$ ($N \geq 2$) and initialize $N$ granularity subpopulations $GS_h$, where $h \in \{1, \ldots, N\}$.
2. Initialize the first granularity subpopulation by assigning the collective preference in round $P_1$ to granularity subpopulation center $GS_1^1$. Then, initialize the second granularity subpopulation center $GS_1^2$ as the elitist preference $E_1^1$, which is the farthest from $E_1^1$.
3. For $GS_h^1 (h \geq 3)$, compute the minimum distance between each of the remaining ordinary elitists’ preferences $E_1^t$ and all current initial granularity subpopulation centers and find the super elitist preference whose minimum distance is the largest by
   \[
   MS(GS_h^1) = \max_x \left( \min_{i<j} d(E_1^t, GS_h^1) \right). \tag{14}
   \]
   Assign it to $GS_h^1$. Repeat this step until all $N$ granularity subpopulations are initialized.
4. In the Granu-subpopulation, to compute the distances between preference relations (both elitists’ preferences and cluster centers indistinctly), the trust degree of different elitists in the same granularity subpopulation is defined by
   \[
   dev(D_t) = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \| P_{ij} - SP_i \|} \tag{15}
   \]
   where $n$ is the total number of elitists, $SP_i$ is the supery elitist, and $P_{ij}$ is the ordinary elitist in the Granu-subpopulation.
5. Compute the trust degree of each super elitist $SP_i$ toward each granularity subpopulation center $GS_h^1 (h \geq 2)$, $u_{GS_h}(P_i) \in [0, 1]$, as follows:
   \[
   u_{GS_h}(P_i) = \frac{1}{\sum_{i=1}^{N} \left( \frac{d(P_i, C_h^1)}{1/d(P_i, C_h^1)} \right)^{1/(b-1)}}. \tag{16}
   \]
6. The granularity subpopulation centers $GS_h^1 (h \geq 2)$ and cluster trust degrees $u_{GS_h}(P_i)$ are updated iteratively. Reputations represent the trust degrees of different granu-subpopulations, which can be expressed as follows:
   \[
   GS_h = \frac{\sum_{i=1}^{m} u_{GS_h}(P_i) P_i}{\sum_{i=1}^{m} u_{GS_h}(P_i)}. \tag{17}
   \]
7. The similarity between granularity subpopulation $GS_h^1 (h \geq 2)$ is determined in the current round $t$, $t \in \{2, \ldots, \text{Maxrounds} - 1\}$, and each granularity subpopulation $C_u^{t-1}(u \geq 2)$ is computed in the previous round $(t-1)$. Thus, the scale of the subdatasets is dynamically updated iteratively by the trust degree relationships based on the subpopulation interactions in different granularity spaces.
8. A granularity subpopulation similarity measure $\text{sim}(GS_h^1, GS_u^{t-1})$ is defined by
   \[
   \text{sim} \left( GS_h^1, GS_u^{t-1} \right) = 1 - \frac{\sum_{i=1}^{m} \Delta_i^t \left( P_i \right)}{m}. \tag{18}
   \]
is greatly improved. Fig. 2 shows the framework of the ensemble consensus compensatory scheme, and its main steps are described in Algorithm 2.

C. Equilibrium Adjustment Strategy of Multigranular Coevolution

Information granularities are not completely independent, and they usually overlap and overlay each other. Therefore, a dynamic-approximation equilibrium adjustment strategy is needed in the multigranular-coevolution space to avoid the super elitists running into optima and, thus, providing good guidance for all elitists. While super elitists are in the local optima state according to the premature judgment mechanism, the mutative-scale equilibrium adjustment strategy is used to ensure a wide range of balancing of exploration and exploitation. This strategy can classify super elitists’ preferences and detect noncooperative behaviors with a global convergence ability and high search accuracy. Fig. 3 shows the updating approximations of the equilibrium adjustment strategy when using different multigranularities for super elitists. Our projection produces different isoceles right triangles, and the arrow shows the direction of sorting super elitists in each triangle area. If two super elitists start with the lower granularity of \( a_{N_j} \), the attribute values of the updating approximations converge to the equilibrium pair \( (a_{N_j}, a_{N_j}) \). Similarly, if both super elitists start with very high granularity of \( a_{N_1} \), the attribute values of the updating approximations converge to the equilibrium pair \( (a_{N_1}, a_{N_1}) \).

Algorithm 1: Continued

where \( \Delta^t_{hu}(P_i) \in [0, 1] \) is the variation in the \( P_i \) membership of both granularity subpopulations, which is computed by

\[
\Delta^t_{hu}(P_i) = \left| \mu^t_{GS_{hu}}(P_i^t) - \mu^{t-1}_{GS_{hu}}(P_i^{t-1}) \right| .
\]  

For the granularity subpopulation \( GS^t_{hu} \), if \( \exists GS^{t-1}_{hu} : \text{sim}(GS^t_{hu}, GS^{t-1}_{hu}) \geq k, k \in [0, 1] \) is a similarity threshold. Then, \( GS^t_{hu} \) and \( GS^{t-1}_{hu} \) are assumed to represent the same granu-subpopulation.

Algorithm 2: Ensemble Consensus Compensatory Scheme (ECCS):

1. Dedicate each granu-subpopulation to its corresponding attribute set. The fitness evaluation is distributed equally among the participating Granu – subpopulation, with the importance weight \( w_i \) of the \( i \)th elitist. Granu – subpopulation, is evaluated in its own domain.

2. Compute the weights by the fuzzy measures over each Granu – subpopulation, \( d_{\sigma(1)} \geq d_{\sigma(2)} \geq \cdots \geq d_{\sigma(n)} \) is the permutation of the evaluation values of the alternative \( A_i \), and \( C_{\sigma(j)} \) is the criterion corresponding to \( d_{\sigma(j)} \). The weight criterion is calculated by

\[
w_{\sigma(j)} = \mu(A_{\sigma(j)}) - \mu(A_{\sigma(j-1)}), \quad j = 1, 2, \ldots, n,
\]

where \( A_{\sigma(j)} = \{C_{\sigma(t)} | t \leq j\} \) for \( j \geq 1, A_{\sigma(0)} = \phi \). \( w_{\sigma(j)} \) is the weight of criterion \( C_{\sigma(j)} \), and \( \mu(A_{\sigma(j)}) \) is the fuzzy measure.

3. Based on the given weight vector \( w = (w_1, w_2, \ldots, w_m)^T \), the intuitionistic fuzzy matrix is a matrix of pairs of nonnegative weight numbers, which is constructed by

\[
I = (a_{ij})_{n \times m}(i = 1, 2, \ldots, n; j = 1, 2, \ldots, m).
\]  

4. Each super elitist \( SP_i \in E \) provides its preference for alternatives according to the fuzzy preference relation \( P_i = (p_{ik})_{n \times n} \), which consists of a matrix of assessments \( p_{ik}^{\ell} \) for each pair of super elitists \( (x_i, x_j), l, k \in \{1, 2, \ldots, n\} \). The ensemble consensus of preferences can be improved if the super elitists provide reciprocal assessments. If \( p_{ik}^l = p, p \in [0, 1], l \neq k \), then \( p_{ik}^l = 1 - p \).

5. Compute the similarity degree of each pair of super elitists \( (SP_i, SP_j) \), and then the similarity matrix \( SM_{ij} = (sm_{ij}^{kl})_{n \times n} \) is defined by

\[
SM_{ij} = \left( \begin{array}{cccc}
sm_{ij}^{11} & \cdots & sm_{ij}^{1n} \\
\vdots & \ddots & \vdots \\
sm_{ij}^{n1} & \cdots & sm_{ij}^{nn}
\end{array} \right) ,
\]  

where \( sm_{ij}^{kl} \) is calculated by

\[
sm_{ij}^{kl} = \frac{1}{n} \sum_{t=1}^{n} \mu_{GS_{ij}}(P_{ij}^{t})^{kl}.
\]

Fig. 3. Coarsening and refining attribute values for the updating approximations in multigranulation space. (a) Dynamic adjustment. (b) Equilibrium.
Algorithm 2: Continued.

where $sm_{ij}^{lk} \in [0, 1]$ is the similarity degree between super elitists $SP_i$ and $SP_j$ in their assessments $p_i^{lk}$ and $p_j^{lk}$, as obtained by the similarity function

$$sm_{ij}^{lk} = 1 - \sqrt{|p_i^{lk} - p_j^{lk}|}.$$  \hfill (23)

6. A consensus matrix $CM = (cm_{ij}^{l})_{n \times m}$ is computed by aggregating similarity matrices, including the importance weights $w_{ij} \in [0, 1]$ of each pair of super elitists ($SP_i$, $SP_j$). The weighted average of the similarity degrees of each elitist $cm_{ik}^{l} \in [0, 1], (l \neq k)$ is computed by

$$cm_{ik}^{l} = \frac{\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} w_{ij} sm_{ij}^{lk}}{\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} w_{ij}}.$$  \hfill (24)

7. If all super elitists are given equal importance weights, $cm_{ik}^{l}$ is redefined by

$$cm_{ik}^{l} = \frac{\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} w_{ij} sm_{ij}^{lk}}{m(m - 1)/2},$$  \hfill (25)

where $m(m - 1)/2$ is the number of different pairs of elitists $(e_i, e_j)$ in the Granu – subpopulation.

8. Design the average weight $cm_{ik}^{l}$ associated with each pair of super elitists $(e_i, e_j)$ to select each pair of features $(x_i, x_k)$ as

$$cm_{ik}^{l} = \frac{w_{ij} sm_{ij}^{lk}}{\sum_{j=i+1}^{m} w_{ij}},$$  \hfill (26)

where $w_{ij} \in [0, 1]$, and the value is computed based on a single super elitist’s weights $w_i, w_j$.

9. An ensemble consensus degree with a compensatory scheme is computed by three different levels:

(i) The level of super elitists $(c_{p}^{lk})$: This value is obtained from $CM$ as $c_{p}^{lk} = cm_{ik}^{l}$, with $l, k \in \{1, \ldots, n\}, l \neq k$.

(ii) The level of super elitists $(c_{a}^{l})$: The level of agreement of each super elitist $x_i \in X$ is computed by

$$c_{a}^{l} = \frac{\sum_{k=1, k \neq l}^{n} c_{p}^{lk}}{n - 1}.$$  \hfill (27)

(iii) The level of preference relation of the overall consensus degree $cr$ is obtained by

$$cr = \frac{\sum_{l=1}^{n} c_{a}^{l}}{n}.$$  \hfill (28)

10. Compare the consensus degree $EM_i$ with a consensus threshold $\mu \in [0, 1]$.

(i) If $EM_i \geq \mu$, the Granu – subpopulation, moves to the attribute selection process.

(ii) If $EM_i < \mu$, the moderator advises the elitists to modify their preferences to increase the level of agreement in the following rounds.

Algorithm 3: Equilibrium Adjustment Strategy of Multigranular Coevolution (EASM):

1. Assuming that $C_h$ and $C_{h}^{-1}$ are considered to be in the same elitist cluster, let $S^{t}_{h} = \sum_{i=1}^{m} \mu_{C_{h}^{i}}(P^{i}_{t})$ and $S^{t-1}_{h} = \sum_{i=1}^{m} \mu_{C_{h}^{i-1}}(P^{i-1}_{t})$ be the sums of the elitists’ membership degrees to cluster $C_{h}^{i}(h \geq 2)$, respectively. Analogously, let $S^{t}_{c} = \sum_{i=1}^{m} \mu_{C_{c}^{i}}(P^{i}_{t})$ be the sum of the elitists’ membership degrees to the collective preference.

2. If $S^{t}_{h} > S^{t-1}_{h}$, the super elitists become more assembled around $C_{h}^{i}$. On the other hand, if $S^{t}_{h} < S^{t-1}_{h}$, the super elitists become less assembled around $C_{1}^{i} = P^{i}_{c}$.

3. Perform a mixed competitive and cooperative convolution among the multigranularity subpopulations and suppose $S_{t}$ as the $t^{th}$ super elitist. For $i = 1$ to $|S_{t}|$, do the following:

(i) Insert the representative of $S_{t}$ as $S_{t, rep}$ into $P^{C}_{t}$.

(ii) If $n_{x} > |S_{t}|$, select the competing super elitist randomly and insert competitors from the selected Granu – subpopulation into $P^{C}_{t}$. Otherwise, insert competitors from other granu-subpopulations into $P^{C}_{t}$ and insert random super elitists from $S_{t}$ into $P^{C}_{t}$.

(iii) Assemble the complete solution with $S_{i,j}$ and the representatives from the other granu-subpopulations, assign Pareto rank to $S_{i,j}$, and calculate niche count of $S_{i,j}$.

(iv) Update the representative super elitist of $S_{t}$ and archive the nondominated solution.

(v) Determine the winning granu-subpopulation $S_{k}$ and update $S_{t} = S_{k}$.

4. Super elitists’ fuzzy membership degrees $\mu_{C_{h}}(P_{t})$ to $Granu – subpopulation$, are computed using similarity measures, where the distance between preference $P_{t}$ and the super elitist center $C_{h}$ is represented as $d(P_{t}, C_{h})$.

5. Calculate the consistency ratio $CR$ for each of the super elitists, $C^{t} = (c_{ij}^{t})_{n \times n}$ and $D^{t} = (d_{ij}^{t})_{n \times n}$, with $t \in \{1, 2, \ldots, s\}$.

approximations convergence to the equilibrium pair $(a_{N_{i}}, a_{N_{i}})$. Thus, the results show that using the equilibrium adjustment strategy of multigranular coevolution for super elitists based on isosceles right triangles can lead to an increase in the size of the basin of multigranularity attraction. Its main steps are described in Algorithm 3.

IV. PROPOSED MSFAR ALGORITHM

We propose a novel MSFAR algorithm to support the formation of hierarchies of information granules of higher types and higher orders to support big data analysis. To accomplish this, we implement the described multigranulation fuzzy–rough sets and supertrust model with a self-evolving compensatory scheme to calculate the multigranular trust degree. It explicitly permits identifying the interdependent variables and adaptively decomposes them in the multigranulation space, so that the complexity
and nonseparability of interdependent variables can be minimized among different fuzzy attribute subsets. It also extracts the fuzzy similarity of the changing knowledge granularity for fuzzy classification with multimodality attributes, effectively solving the problem of missing data in incomplete large-scale information systems. The proposed MSFAR algorithm can split a large data set into reasonable subdatasets with the multigranulation supertrust model. It incorporates an additional multigranulation module and supertrust coevolution to achieve the desirable goal of detecting complex interdependent variables, which can serve as a guide to carry out fuzzy–rough attribute reduction tasks with multigranulation flexible classification thresholds in big data. Its main steps are detailed as follows.

First, the fuzzy attribute sets are mapped into the evolutionary population space, and the fuzzy reduction model is completed as the optimization objective model.

Second, a multigranulation fuzzy–rough model based on the valued tolerance relation is constructed to identify the fuzzy similarity of the changing knowledge granularity for fuzzy classification with multimodality attributes, which effectively solves the problem of missing data in an incomplete large-scale information system. Then, an equilibrium method of multigranular coevolution is employed to classify super elitists’ preferences and detect noncooperative behaviors.

Third, the multigranulation supertrust-coevolution model with a self-evolving compensatory scheme is adopted to calculate the multigranular trust degrees of different granularities, which represent their reputations in the group, resulting in the creation of reasonable subproblems with different levels from coarsened to refined granulation. It can self-adapt among different multigranulation layers and capture interdependent fuzzy–rough attribute subsets.

The pseudocodes of MSFAR are listed in Algorithm 4.

V. EXPERIMENTAL EVALUATION AND DISCUSSIONS

To validate the efficiency and effectiveness of the proposed MSFAR algorithm, we carry out a thorough series of experiments in this section, including a comparison of the computational times and accuracies of different algorithms and robustness comparisons for big datasets with consideration of the attribute noise. We provide performance comparisons between MSFAR and other representative algorithms for large-scale datasets and conduct robustness comparisons using attribute noise datasets for MSFAR versus representative algorithms.

A. Experimental Setup

All experiments are performed on computers with a Windows 10 operating system, an i7 3770k Intel CPU and 64 GB RAM using Java 11 programming language. We select five publicly available large-scale datasets from the UCI repository with heterogeneous attributes [46]. In addition, we produce three synthetic large-scale WEKA datasets with the WEKA data mining software [47], with these datasets involving a large number of samples with different statistical characteristics. Descriptive information about the eight datasets is given in Table I, where “KddCup99” is the network connectivity data set from USA air

| Algorithm 4: Proposed MSFAR algorithm. |
|----------------------------------------|
| 1. Initialize the searching space of the fuzzy attribute sets and the granularity subpopulations by assigning the collective preference in round \( P_t \), initialize the granularity subpopulations \( GS_h \), and generate a list of candidate fuzzy–rough attribute subsets \( \{A_1, A_2, \ldots, A_n\} \). |
| 2. Decompose the fuzzy–rough attribute sets, compute equivalence class of the decision table, and classify super elitists’ preferences using Algorithm 2 (ECCS) to determine whether a granularity subpopulation is composed of the same super elitists as follows. Suppose that \( GS_h^t \) and \( GS_h^{t-1} \) are considered to represent the two granularity subpopulations of two super elitists and ordinary elitists and that their super elitists’ trust degrees \( \mu_{GS_h^t}(P_i^t) \) and \( \mu_{GS_h^{t-1}}(P_i^{t-1}) \) have values close to each other for all \( e_i \in E \). Then, conduct two similar granularity subpopulation compositions. |
| 3. Conduct the ensemble consensus compensatory scheme by using Algorithm 2 (ECCS) and then obtain the corresponding perfect consistency equilibrium degree \( CED \) based on isosceles right triangles as \( \tilde{C}^t = (\tilde{c}_{ij}^t)_{n \times n} \) for any inconsistent \( CED \) \( C^t = (c_{ij}^t)_{n \times n} \), where \( \tilde{c}_{ij}^t = \frac{c_{ij}^t}{2} \) for \( i, j = 1, 2, \ldots, n \). |
| 4. Perform the equilibrium adjustment strategy of multigranular coevolution using Algorithm 3 (EASMFC) and then obtain the corresponding perfect consistency equilibrium degree \( CED \) based on isosceles right triangles as \( \tilde{C}^t = (\tilde{c}_{ij}^t)_{n \times n} \) for any inconsistent \( CED \) \( C^t = (c_{ij}^t)_{n \times n} \), where \( \tilde{c}_{ij}^t = \frac{c_{ij}^t}{2} \) for \( i, j = 1, 2, \ldots, n \). |
| 5. Obtain the perfect CED as \( \tilde{D}^t = (\tilde{d}_{ij}^t)_{n \times n} \) for \( t \in \{1, 2, \ldots, s\} \). Thus, a perfect consistency pair \( \tilde{C}^t = (\tilde{C}^t, \tilde{D}^t) \) can be constituted by \( \tilde{C}^t = (\tilde{c}_{ij}^t)_{n \times n} \) and \( \tilde{D}^t = (\tilde{d}_{ij}^t)_{n \times n} \) for \( t \in \{1, 2, \ldots, s\} \). |
| 6. The energy function is reconstructed by |
| \[ EF(e_k|p_{N_i}^{(m)}) = \sum_{j=1}^{K} G_{kj} v_j, \] |
| where \( G \) denotes a \( K \times K \) matrix defining the connectivity between elitist populations \( k \) and \( j \) and \( v_j \) denotes the accumulated class probabilities in the neighborhood information \( N_i \) of elitist population \( i \). |
| 7. The definition of the multigranulation flexible threshold \( FT_R(X) \) is formulated as |
| \[ FT_R(X) = \frac{|B_{nR}(X)|}{|U|} \log_2|U| - \sum_{i=1}^{n} \frac{|X_i|}{|U|} \log_2 \left( \frac{1}{|X_i|} \right), \] |
| where \( B_{nR}(X) \) is the threshold region of the boundaries. |
| 8. Conduct classification and learning of fuzzy–rough attribute sets with a parallel accelerating framework and \( FT_R(X) \) as follows: |
| (i) Calculate the degree of dependency as the criterion for attribute selection and obtain the similarity degree \( VS_{(a)}(x, y) \) based on quantitative tolerance. |
| (ii) Select feature subsets \{subset_1, \ldots, subset_i, \ldots, subset_n\}. |
Algorithm 4: Continued

(iii) Select the best feature subset $FS_i^{best}$ for each granu-subpopulation and achieve a cascade feature set of fuzzy–rough attribute subsets as follows:

$$FS = FS_1^{best} \cup FS_2^{best}, \ldots, \cup FS_n^{best}.$$  \hfill (32)

9. Evaluate whether or not the accuracy of the fuzzy–rough attribute reduction is satisfied with respect to the predefined accuracy. If satisfied, then output the optimal set $FS_{Opt} = \bigcup_{i=1}^{n} FS_i^{best}$; otherwise, go to Step 6.

| No | Datasets | Samples | Attributes |
|----|----------|---------|------------|
| 1  | PokerHand| 1,025,010 | 10         |
| 2  | KddCup99 | 5,000,000 | 41         |
| 3  | Susy     | 5,000,000 | 18         |
| 4  | RLCP     | 5,749,132 | 14         |
| 5  | Higgs    | 11,000,000| 28         |
| 6  | Weka-1.8G| 32,000,000| 10         |
| 7  | Weka-3.2G| 40,000,000| 15         |
| 8  | Weka-6.4G| 80,000,000| 15         |

In the following experiment, we further evaluate the classification accuracy of MSFAR for selected dynamically increasing sample sizes in the large-scale Higgs and Weka-1.8G datasets compared with those of representative algorithms. We employ only two classifiers, namely, support vector machine (SVM) [52] and C4.5 [53], to process datasets whose attributes have been selected by six different methods, namely, B-FRFS, MIBARFRAR, UFRFS, MLFRS, and MFDAR. The attribute reduction time denotes the running time required for different fuzzy attribute reduction algorithms select features. For example, in the Weka-6.4G data set, the attribute reduction time of MSFAR is 2728 s, whereas the corresponding times for B-FRFS, MIBARFRAR, UFRFS, MLFRS, and MFDAR are 4916 s, 6321 s, 5129 s, 4723 s and 3643 s, respectively. The computational time of MSFAR is only 55.49%, 43.16%, 53.19%, 57.76%, and 74.88% of those of B-FRFS, MIBARFRAR, UFRFS, MLFRS, and MFDAR, respectively. Similar results are evident for most datasets. Table III shows the numerical comparison of the selected features of different fuzzy attribute reduction algorithms. As described in Tables II and III, the proposed MSFAR algorithm can reduce the amount of redundant uncertain, unstructured, and imprecise data and significantly improve the computational time. Therefore, it acquires the optimal number of selected features of the fuzzy–rough attribute sets, which is consistently a much better performance than those of its rivals for seven large-scale datasets. The main reasons behind these results are that the multigranulation supertrust-coevolution model with a self-evolving compensatory scheme employed in MSFAR is constructed to calculate the multigranular trust degree at different granularities and to split large datasets into reasonable subdatasets. MSFAR can consider both strongly relevant features and their corresponding correlated features simultaneously, and it selects important correlated features from a set of attribute features for classification.

B. Computational Times and Accuracies of Different Algorithms

To test the computational feasibility of MSFAR for use with large-scale datasets, we quantitatively compare its classification accuracy with those of some representative fuzzy attribute reduction algorithms, such as fuzzy boundary-region-based FS (B-FRFS) [48], mutual information-based algorithm for fuzzy-rough attribute reduction (MIBARFRAR) [36], unsupervised fuzzy-rough feature selection (UFRFS) [49], multimodality fuzzy data attribute reduction (MFDAR) [50] and multi-label fuzzy rough set (MLFRS) [51]. To avoid the influence of random selection, independent runs for each data set are repeated 50 times, and the averaged computational time and accuracy are presented as the final results. As described in Table II, we employ “-” to indicate no acceptable solution and a bold number to represent the best result of the computational time (time/s $\times 10^5$). As indicated in Table II, MSFAR clearly exhibits the highest speed of fuzzy–rough attribute reduction for most datasets compared with those of B-FRFS, MIBARFRAR, UFRFS, MLFRS and MFDAR. The attribute reduction time denotes the running time required for different fuzzy attribute reduction algorithms select features. For example, in the Weka-1.8G data set, the attribute reduction time of MSFAR is 2728 s, whereas the corresponding times for B-FRFS, MIBARFRAR, UFRFS, MLFRS, and MFDAR are 4916 s, 6321 s, 5129 s, 4723 s and 3643 s, respectively. The computational time of MSFAR is only 55.49%, 43.16%, 53.19%, 57.76%, and 74.88% of those of B-FRFS, MIBARFRAR, UFRFS, MLFRS, and MFDAR, respectively. Similar results are evident for most datasets. Table III shows the numerical comparison of the selected features of different fuzzy attribute reduction algorithms. As described in Tables II and III, the proposed MSFAR algorithm can reduce the amount of redundant uncertain, unstructured, and imprecise data and significantly improve the computational time. Therefore, it acquires the optimal number of selected features of the fuzzy–rough attribute sets, which is consistently a much better performance than those of its rivals for seven large-scale datasets. The main reasons behind these results are that the multigranulation supertrust-coevolution model with a self-evolving compensatory scheme employed in MSFAR is constructed to calculate the multigranular trust degree at different granularities and to split large datasets into reasonable subdatasets. MSFAR can consider both strongly relevant features and their corresponding correlated features simultaneously, and it selects important correlated features from a set of attribute features for classification.

In the following experiment, we further evaluate the classification accuracy of MSFAR for selected dynamically increasing sample sizes in the large-scale Higgs and Weka-1.8G datasets compared with those of representative algorithms. We employ only two classifiers, namely, support vector machine (SVM) [52] and C4.5 [53], to process datasets whose attributes have been selected by six different methods, namely, B-FRFS, MIBARFRAR, UFRFS, MLFRS, and MFDAR. Tables IV and V report the classification accuracy versus the dynamically increasing sample size of large datasets with the SVM classifier and C4.5 classifier, respectively. The six different attribute reduction algorithms result in the different sets of attributes for the large-scale increasing datasets. It is obvious that MSFAR significantly surpasses most of the representative algorithms. As an extreme case, both B-FRFS and MIBARFRAR fail with increasing sample sizes of the Higgs data set because they are easily overwhelmed when processing high-dimensional large datasets. However, MSFAR achieves much better classification accuracy because it can benefit from the advantages of that the multigranulation fuzzy–rough set model can accurately capture interdependent variables associated with structurally complex and incomplete attribute sets and can greatly eliminate most irrelevant attribute sets without lowering the classification performance. For example, for the $80 \times 10^6$ sample size of Weka-1.8G, MSFAR can reach 97.81% classification accuracy, whereas for B-FRFS, UFRFS, and MLFRS, these values are 92.21%, 93.43%, and 94.62%, respectively. For the evergrowing large-scale datasets, MSFAR performs significantly better. Therefore,
the performance of MSFAR improves for large datasets—the larger is the data set, the higher is the classification accuracy. As indicated in experimental results, MSFAR is suitable for use in fuzzy–rough attribute reduction and classification for large-scale increasing datasets, thereby overcoming the limitations of the representative algorithms.

The experimental results clearly indicate that the classification system employing MSFAR as the fuzzy–rough attribute reduction algorithm can acquire the optimal reduction results of structurally complex and incomplete attribute sets and lead to an appealing performance in the classification accuracy, irrespective of different classifiers.

C. Statistical Analysis

In order to sufficiently report classification accuracies, an appropriate statistical test need to be applied to evaluate the statistical significance of the results of the fuzzy–rough attribute reduction algorithms. In this article, a paired \( t \)-test was used to determine the statistical significance of the results at the 0.05 level.

### TABLE II

**COMPUTATIONAL TIME (IN SECONDS × 10^2) COMPARISON OF DIFFERENT FUZZY ATTRIBUTE REDUCTION ALGORITHMS**

| Datasets   | B-FRFS | MIBARFRAR | UFRFS | MLFRS | MFDAR | MSFAR |
|------------|--------|-----------|-------|-------|-------|-------|
| 1. PokerHand | 29.61±0.46 | 27.89±1.41 | 32.45±0.52 | 33.09±1.29 | 26.34±0.20 | 21.35±0.23 |
| 2. KiddCup99 | 45.18±1.28 | 40.98±0.35 | 36.89±0.38 | 41.51±1.41 | 34.90±0.62 | 24.23±0.24 |
| 3. SySy     | 34.61±1.41 | 38.02±1.48 | 39.36±2.51 | 34.07±0.32 | 30.12±0.23 | 27.09±0.20 |
| 4. R e p l  | 35.92±0.37 | —          | 29.35±1.32 | 31.10±0.28 | 32.24±0.45 | 23.97±0.31 |
| 5. Higgs    | —        | 30.91±0.42 | 20.51±0.34 | 19.25±0.41 | 23.21±0.32 | 15.33±0.38 |
| 6. Weka-1.8G| 41.19±0.53 | —          | 47.29±1.37 | 34.22±0.31 | 35.32±0.87 | 32.23±0.28 |
| 7. Weka-3.2G| 31.68±2.47 | 38.11±0.68 | 23.90±1.51 | 24.19±0.48 | 24.32±0.39 | 19.56±0.39 |
| 8. Weka-6.4G| 49.16±1.54 | 63.21±2.59 | 51.29±1.48 | 47.23±0.37 | 36.43±0.69 | 27.28±0.40 |

### TABLE III

**NUMERICAL COMPARISON OF THE SELECTED FEATURES OF DIFFERENT FUZZY ATTRIBUTE REDUCTION ALGORITHMS**

| Datasets   | B-FRFS | MIBARFRAR | UFRFS | MLFRS | MFDAR | MSFAR |
|------------|--------|-----------|-------|-------|-------|-------|
| 1. PokerHand | 6      | 6         | 7     | 8     | 8     | 9     |
| 2. KiddCup99 | 32     | 34        | 31    | 36    | 37    | 40    |
| 3. SySy     | 13     | 14        | 14    | 15    | 16    | 18    |
| 4. R e p l  | 12     | —         | 11    | 12    | 13    | 14    |
| 5. Higgs    | —      | 18        | 22    | 24    | 28    | 25    |
| 6. Weka-1.8G| 7      | —         | 6     | 7     | 8     | 8     |
| 7. Weka-3.2G| 12     | 12        | 13    | 12    | 14    | 15    |
| 8. Weka-6.4G| 12     | 11        | 13    | 11    | 12    | 15    |

### TABLE IV

**CLASSIFICATION ACCURACY COMPARISON WITH THE SVM CLASSIFIER FOR THE INCREASING LARGE-SCALE EVERGROWING HIGGS DATA SETS (TEST±STD/%)**

| Algorithms  | Sample size (× 10^8) |
|-------------|---------------------|
|             | 2                   |
|             | 4                   |
|             | 10                  |
|             | 20                  |
|             | 40                  |
| MSFAR       | 91.13±0.23*         |
| B-FRFS      | 87.89±0.31          |
| MIBARFRAR   | 82.19±0.28          |
| UFRFS       | 87.12±0.56          |
| MLFRS       | 88.21±0.28          |
| MFDAR       | 89.23±0.23          |
| SVM         | 93.21±0.35          |
|            | 95.18±0.41          |
|            | 96.24±0.29          |
|            | 96.93±0.38          |
|            | 89.29±0.51          |
|            | 84.39±0.42          |
|            | 91.29±0.27          |
|            | 89.36±0.76          |
|            | 91.47±0.49          |
|            | 91.30±0.43          |
|            | 92.23±0.54          |
|            | 93.10±0.31          |
|            | 92.54±0.36          |
|            | 92.45±0.67          |
|            | 93.78±0.34          |
level. Since a paired $t$-test is a parametric test, it always assumes approximate normality. The results are shown in Table VI for the SVM classifier, where a “b” symbol next to a value indicates that the performance is statistically better than MSFAR, and a “w” symbol shows that the performance is worse statistically. The final line in Table VI summarizes the count of the number of statistically better, equivalent, and worse results for each representative fuzzy attribute reduction algorithm in comparison with MSFAR. The statistical comparison results between each compared algorithm and MSFAR indicate that only three datasets (KddCup99, Susy, and RLCP) for which MSFAR is bettered by MLFRS and MFDAR, but for the remainder, MSFAR achieves statistically equivalent to or better accuracy results than other representative fuzzy attribute reduction algorithms.

It is obvious in Table VI that B-FRFS and MIBARFRAR do not perform as well as does MSFAR in producing classification accuracy results, which are consistent across the SVM classifier, as there are four or five cases wherein they produces worse statistical results compared with MSFAR. From the results reported in this table, it can be seen that the representative fuzzy attribute reduction algorithms attain a few results that are statistically comparable with the accuracy for the unreduced data, but MSFAR achieves significantly better classification results than them. Generally speaking, it can be concluded that MSFAR is superior to the other algorithms in the paired $t$-test. It offers the significance level for the tests, whereas retaining the semantics of the data. The better performance of the proposed MSFAR algorithm is achieved due to the fact that it provides an efficient way to select a reduced attribute set of real-valued datasets, having maximum significance and relevance without lowering the classification performance.

### D. Discussion

From the results for average classification accuracy presented in Tables IV and V, it can be observed that the multigranulation fuzzy–rough set model can perform better with lower average subset sizes in comparison with both the UFRFS and MLFRS algorithms. In particular, for the proposed MSFAR algorithm, from the abovementioned results, it is clear that the multigranulation model offers a greater reduction of the large-scale size. This reduction is to be expected since there is much discriminative information contained in the decision features. The main reason behind this result is that the multigranulation fuzzy–rough set
model can accurately capture interdependent variables associated with structurally complex and incomplete attribute sets and can greatly eliminate most irrelevant attribute sets without lowering the classification performance. The possible reasons behind this effect are that we construct a multigranulation fuzzy–rough set model based on a valued tolerance relation to extract the fuzzy similarity of the changing knowledge granularity for fuzzy classification with consideration of multimodality attributes, which effectively solves the problem of missing data in the large-scale information system. Meanwhile, the multilayered supertrust-coevolution model with a self-evolving compensatory scheme can calculate the multigranular trust degree at different granularities based on the reputation in the group to split the large problem into reasonable subproblems.

It is necessary for the algorithm to perform efficiently in real time so that it consistently functions without requiring a significant amount of computing resources. In our experiments, the time complexity of the proposed MSFAR algorithm is also analyzed by observing the real-time performance for selected big datasets of varying sizes. The experimental results indicate that the computational complexity of MSFAR is obviously less than those of UFRFS and MLFRS. Hence, if we were to perform the evaluation using larger datasets, the computational time would be unacceptable for the comparison algorithms, but MSFAR requires a smaller amount time to obtain the optimal solution.

Of course, in a few special cases, the results of MSFAR are slightly poorer than those of the representative algorithms. In general, the two main components in MSFAR should work together to allow it to obtain better results. Furthermore, MSFAR can dynamically adapt its main operators to suit various large-scale instances with dynamically increasing percentages of noise.

VI. CONCLUSION

Recently, big data has been an emerging topic that has attracted the attention of many researchers. The significant amount of unstructured, uncertain, and imprecise large-scale data exhibits structurally complex and granular characteristics. Uncertainty data have been widely adopted for attribute reduction, but alone, they may be insufficient for use in batch feature selection. The recent progress in fuzzy–rough set approaches can be helpful for analyses of big data problems. In this article, we present a novel MSFAR algorithm for use in big data analysis at different granularity levels. A multigranulation fuzzy–rough set model based on a valued tolerance relation is constructed to identify the fuzzy similarity of the changing knowledge granularity for fuzzy classification with multimodality attributes. Meanwhile, the multigranulation supertrust-coevolution model with a self-evolving compensatory scheme was adopted to calculate the multigranular trust degree at different granularities, and it could be directly applied to a variety of knowledge analytical problems with continuous or numerical large-scale datasets. The experimental results demonstrate that MSFAR produces very good results. It is theoretically and experimentally indicated that the multigranulation supertrust-coevolution model with a self-evolving compensatory scheme could provide a much better performance by the MSFAR compared with those of the representative models. These represent important developments for improving the reasoning and understanding of big data.

In the era of big data, the size of large data usually dynamically increases, including current changing and interconnected datasets. It is time-consuming to perform efficient attribute reduction and classification for these uncertain and redundant datasets. In the future, it is expected that using analytical methods to learn from big data can significantly improve the fuzzy–rough attribute reduction process. We will also explore the effective and robust multigranulation mechanisms of fuzzy–rough reduction estimation to achieve improved understanding of large-scale feature selection. We intend to exert great effort in promoting our research to offer a new avenue by which to address the problem of optimum predicting disorder from neonatal brain magnetic resonance imagings.

ACKNOWLEDGMENT

The authors would like to express the sincere appreciation to the anonymous reviewers for their insightful comments that greatly improve the quality of this article.

REFERENCES

[1] C.-H. Fung, Z. T. H. Tse, and K.-W. Fu, “Converting big data into public health,” *Science*, vol. 347, no. 6222, Art. no. 620.
[2] C. Min, M. Shiwen, and L. Yunhao, “Big data: A survey,” *Mobile Netw. Appl.*, no. 2, pp. 171–209, Apr. 2014.
[3] C. P. Chen and C. Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on big data,” *Inf. Sci.*, vol. 275, pp. 314–347, Aug. 2014.
[4] H. V. Jagadish et al., “Big data and its technical challenges,” *Commun. ACM*, vol. 57, no. 7, pp. 86–94, Jul. 2014.
[5] A. Phinyomark, E. Ibáñez-Marcelo, and G. Petr, “Resting-state fMRI functional connectivity: Big data preprocessing pipelines and topological data analysis,” *IEEE Trans. Big Data.*, vol. 3, no. 4, pp. 415–428, Oct.–Dec. 2017.
[6] L. Dong et al., “A hierarchical distributed processing framework for big image data,” *IEEE Trans. Big Data.*, vol. 2, no. 4, pp. 297–309, Oct.–Dec. 2016.
[7] M.-P. Hosseini, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh, “Optimized deep learning for EEG big data and seizure prediction BCI via internet of things,” *IEEE Trans. Big Data.*, vol. 3, no. 4, pp. 392–404, Oct.–Dec. 2017.
[8] A. Segatori, A. Bechini, P. Ducange, and F. Marcelloni, “A distributed fuzzy associative classifier for big data,” *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2656–2669, Sep. 2018.
[9] S. Madden, “From databases to big data,” *IEEE Internet Comput.*, vol. 16, no. 3, pp. 4–6, May–Jun. 2012.
[10] J. Li and H. Liu, “Challenges of feature selection for big data analytics,” *IEEE Intell. Syst.*, vol. 32, no. 2, pp. 9–15, Mar.–Apr. 2017.
[11] M. Duan, K. Li, and X. Liao, “A parallel multiclassification algorithm for big data using an extreme learning machine,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2337–2351, Jun. 2018.
[12] L. A. Zadeh, “Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic,” *Fuzzy Sets Syst.*, vol. 90, no. 2, pp. 111–127, 1997.
[13] W. Pedrycz, *Granular Computing: Analysis and Design of Intelligent Systems*. Boca Raton, FL, USA: CRC Press, 2013.
[14] L. A. Zadeh, “Fuzzy sets,” *Inf. Control*, vol. 8, pp. 338–353, Jun. 1965.
[15] Z. Pawlak, “Rough sets,” *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, Oct. 1982.
[16] Z. Bonikowski and E. Brynarski, “Extensions and intensions in rough set theory,” *Inf. Sci.*, vol. 107, no. 1–4, pp. 149–167, Jun. 1998.
Witold Pedrycz (Fellow, IEEE) received the M.Sc. degree in computer science in 1977, the Ph.D. degree in computer engineering in 1980, and the D.Sc. degree in system science in 1984, all from the Silesian University of Technology, Gliwice, Poland.

He is currently a Professor and Canada Research Chair in Computational Intelligence with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland. He holds an appointment of special professorship with the School of Computer Science, University of Nottingham, Nottingham, U.K. He has published numerous papers in the above-mentioned areas. He has also authored 15 research monographs. His current research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy control, and software engineering.

Dr. Pedrycz was the recipient of the prestigious Norbert Wiener Award from the IEEE Systems, Man, and Cybernetics Society in 2007, IEEE Canada Computer Engineering Medal 2008, Cajastur Prize for Soft Computing from the European Centre for Soft Computing for “pioneering and multifaceted contributions to granular computing” in 2009, Killam Prize in 2013, and the Fuzzy Pioneer Award 2013 from IEEE Computational Intelligence Society. He was elected as a Foreign Member of the Polish Academy of Sciences in 2009 and a Fellow of the Royal Society of Canada in 2012. He has been a Member of numerous program committees of the IEEE conferences in the area of fuzzy sets and neurocomputing. He is intensively involved in editorial activities. He is the Editor-in-Chief of Information Sciences, WIREs Data Mining and Knowledge Discovery (Wiley), and International Journal of Granular Computing (Springer). He serves on an Advisory Board of the IEEE TRANSACTIONS ON FUZZY SYSTEMS and is a Member of a number of editorial boards of other international journals.

Isaac Triguero (Member, IEEE) received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Granada, Spain, in 2009 and 2014, respectively.

He has been an Assistant Professor of data science since June 2016. His work is mostly concerned with the research of novel methodologies for big data analytics. He has published more than 65 international publications in the fields of big data, machine learning and optimization. He has an h-index = 21 and more than 1700 citations on Google Scholar.

Dr. Triguero is a Section Editor-in-Chief of the Machine Learning and Knowledge Extraction, and an Associate Editor for the Big Data and Cognitive Computing, and the IEEE ACCESS. He has acted as the Program Co-Chair of the IEEE Conference on Smart Data in 2016), the IEEE Conference on Big Data Science and Engineering in 2017, and the IEEE International Congress on Big Data in 2018. He is currently leading a KTP fund project funded by Innovative UK and the energy provider E.ON that investigates smart metering data.

Zehong Cao (Member, IEEE) received the and B.S. degree from Northeastern University, in 2012, the M.S. degree from The Chinese University of Hong Kong, in 2013, and the dual Ph.D. degree in information technology from the University of Technology Sydney (UTS), Australia, and in electrical and control engineering from National Chiao Tung University, Taiwan, in 2017.

He is currently an Assistant Professor with the Discipline of Information and Communication Technology, School of Technology, Environments and Design, College of Sciences and Engineering, University of Tasmania, Hobart, TAS, Australia, and an Adjut Fellow with the Faculty of Engineering and IT, UTS. He has 30+ publications in the most respected journals. His current research interests include fuzzy sets and systems, fuzzy neural networks, game artificial intelligence, deep reinforcement learning, brain–computer interface, and biosignal processing.

Dr. Cao serves as an Associate Editor for the IEEE ACCESS and the Journal of Intelligent and Fuzzy Systems and a Guest Editor for Swarm and Evolutionary Computation, Neurocomputing, and International Journal of Distributed Sensor Networks. He was the recipient of the UTS Centre for Artificial Intelligence Best Paper in 2017, UTS Faculty of Engineering and IT Publication Award in 2017, UTS President Scholarship in 2015, and NCTU & Songshanhu Scholarship in 2014.

Chin-Teng Lin (Fellow, IEEE) received the B.S. degree in computer engineering from National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 1986, and the M.S. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1989 and 1992, respectively.

He is currently a Distinguished Professor with the Faculty of Engineering and Information Technology, University of Technology Sydney, Australia, where he is also the Co-Director of the Center for Artificial Intelligence. He is also an Honorary Chair Professor of electrical and computer engineering with NCTU and an Honorary Professor with the University of Nottingham. He has coauthored Neural Fuzzy Systems (Prentice-Hall, 1996) and authored Neural Fuzzy Control Systems With Structure and Parameter Learning (World Scientific, 1994). He has published more than 300 journal papers, with a total of 19,232, h-index of 64, 110-index of 243) in the areas of neural networks, fuzzy systems, brain–computer interface, multimedia information processing, and cognitive neuroengineering, including more than 120 IEEE journal papers.

Dr. Lin was elevated to International Fuzzy Systems Association Fellow in 2012. He was the recipient of the IEEE Fuzzy Systems Pioneer Awards in 2017. He was the Editor-in-Chief for the IEEE TRANSACTIONS ON FUZZY SYSTEMS from 2011 to 2016. He also served on the Board of Governors at IEEE Circuits and Systems Society from 2005–2008, the IEEE Systems, Man, Cybernetics Society during 2003–2005, and the IEEE Computational Intelligence Society during 2008–2010. He was the Distinguished Lecturer of the IEEE CAS Society from 2003 to 2005 and CIS Society from 2015 to 2017. He was the Chair of the IEEE-CIS Distinguished Lecturer Program Committee in 2018. He served as the Deputy Editor-in-Chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II during 2006–2008.