

Co-evolutionary Fuzzy Attribute Order Reduction With Complete Attribute-value Space Tree

Weiping Ding, *Member, IEEE*, Isaac Triguero, and Chin-Teng Lin, *Fellow, IEEE*

Abstract— Since big data sets are structurally complex, high-dimensional, and their attributes exhibit some redundant and irrelevant information, the selection, evaluation, and combination of those large-scale attributes pose huge challenges to traditional methods. Fuzzy rough sets have emerged as a powerful vehicle to deal with uncertain and fuzzy attributes in big data problems that involve a very large number of variables to be analyzed in a very short time. In order to further overcome the inefficiency of traditional algorithms in the uncertain and fuzzy big data, we present a new co-evolutionary fuzzy attribute order reduction algorithm (CFAOR) based on a complete attribute-value space tree. A complete attribute-value space tree model of decision table is designed in the attribute space to adaptively prune and optimize the attribute order tree. The fuzzy similarity of multi-modality attributes can be extracted to satisfy the needs of users with the better convergence speed and classification performance. Then the decision rule sets generate a series of rule chains to form an efficient cascade attribute order reduction and classification with a rough entropy threshold. Finally, the performance of CFAOR is assessed with a set of benchmark problems that contain complex high dimensional datasets with noise. The experimental results demonstrate that CFAOR can achieve the higher average computational efficiency and classification accuracy, compared with the state-of-the-art methods. Furthermore, CFAOR is applied to extract different tissues surfaces of dynamical changing infant cerebral cortex and it achieves a satisfying consistency with those of medical experts, which shows its potential significance for the disorder prediction of infant cerebrum.

Index Terms— Complete attribute-value space tree, Co-evolutionary fuzzy attribute order reduction, Rough entropy threshold, Tissues extraction of infant cerebral cortex

I. INTRODUCTION

In recent years, with the development of various technologies, a large number of data are being continually generated around us. Big data have attracted plenty of attention from a

variety of fields such as biology, health, business management, cognition analysis in human brain, and so on [1]-[3]. The data contents include a complex mixture of texts, speeches, images, and videos [4][5]. It is also true that massive amount of data can potentially provide a much deeper understanding of both nature and society, opening up new ways for research. It is, however, a challenging task to extract useful knowledge from such big data.

It has been observed that many real datasets are usually structurally complex, high-dimensional, and multi-granular. Those attributes usually exhibit some irrelevant and redundancy information. In these cases, the big datasets increase dynamically in size, which occur in a few of fields such as the public health and welfare, economics analysis, and medical research. Therefore, a series of emerging topics such as big data acquisition, storage, management and processing are important issues [6][7]. It becomes highly desirable to develop some effective representative methodologies to analyze big data and further handle their characteristics, such as redundancy, uncertainty, fuzzy, and heterogeneity. The massive amount of data makes traditional data analytical methods inadequate to tackle many real-world high dimensional problems. Although a large number of candidate attribute sets is provided, most of them may turn out to be redundant or irrelevant, which heavily deteriorates the performance of traditional methods.

With the over-flooding of big data, researchers and practitioners have started showing remarkable interest to explore the data space, and have considered that structuralized knowledge reasoning is an effective computational paradigm for dealing with big data tasks. Granular computing (GC) focuses on the knowledge representation and reasoning with information granules, and fuzzy sets and rough sets are two crucial branches of GC [8][9]. Fuzzy set theory (FST) was introduced by Zadeh in 1965 to represent concepts with ambiguous boundaries and to understand the processes of complex human reasoning [10]. It has become a popular tool for the design of fuzzy classifiers. However, a fuzzy set is only characterized by a membership function, which largely ignores the uncertainty. Rough set theory (RST) was presented by Pawlak in 1982 to quantitatively analyze the uncertainty and to process incomplete knowledge [11]. It can find a decision-making table between the strict statistics and random distribution. Since RST can typically describe the uncertainty of knowledge, it has been extensively used in data mining, knowledge discovery, and intelligent system [12]-[16]. Fuzzy rough sets (FRS) appear as a combination of the advantages of two complementary areas (RST and FST), which provides an effective way to overcome the problem of discretization. It can be widely applied to various kinds of attribute reduction problems of numerical or continuous large-scale datasets [17]-[20]. FRS is defined by two fuzzy sets, fuzzy lower and

* This work was supported in part by the Australian Research Council (ARC) under discovery grant DP180100656 and DP180100670, the National Natural Science Foundation of China under Grant 61300167, Six Talent Peaks Project of Jiangsu Province under Grant XYDXXJS-048, Jiangsu Provincial Government Scholarship Program under Grant JS-2016-065, Natural Science Foundation of Jiangsu Province under Grant BK20151274, and Applied Basic Research Program of Nantong under Grant GY12016014. (*Corresponding author: Weiping Ding*)

W. Ding is with the School of Computer Science and Technology, Nantong University, Nantong 226019, China (e-mail: dwp9988@163.com)

I. Triguero is with the School of Computer Science, University of Nottingham, United Kingdom (e-mail: Isaac.Triguero@nottingham.ac.uk)

C.-T. Lin is with the Centre for Artificial Intelligence, FEIT, University of Technology Sydney, Ultimo NSW 2007, Australia (e-mail: chin-teng.lin@uts.edu.au)

upper approximations, which are obtained by extending the corresponding crisp rough set notions. Elements that belong to the lower approximation are considered to belong to the approximated set with absolute certainty. Elements in the fuzzy rough case have a membership in the range, which allows for a greater flexibility in handling uncertain information [21] [22]. Thus, there is a good potential to improve reasoning and understanding of big data by using a FRS method.

A. Related Work

In recent years, some significant algorithms and models based on FRS have been presented. Zhao *et al.* [23] used one generalized FRS model to construct a rule-based classifier, in which the consistence degree was used for the reasonable critical value to keep the discernibility information invariant. The experimental results showed this model was effective and feasible on noisy data, whereas, its computational capability in big data needed to further be improved. Jensen and Cornelis [24] exploited the concepts of lower and upper approximations based on FRS and put forward a new nearest neighborhood algorithm to classify all test objects and predict their decision values. Experimental results showed that this algorithm was competitive with some leading classification methods. However, one obvious limitation was that no way was designed to handle the data possessing missing values. Hu *et al.* [25] summarized the properties of typical fuzzy rough models in handling noisy tasks and revealed why they were sensitive to the level of noise on fuzzy rough computation. Then a collection of robust FRS models based on fuzzy lower approximations is developed, and the experiments results on real-world tasks illustrated the effectiveness of these models. Parthaláin and Jensen [26] used FRS to select features for inclusion and removal from the final candidate subset and presented two unsupervised feature selection approaches as UFRFS and dUFRFS. The approaches were shown to retain useful features. But UFRFS utilized a simple but nevertheless effective backwards elimination way for search, whilst dUFRFS adopted a greedy forward selection way. Furthermore, two search techniques often returned sub-optimal results. Zeng *et al.* [27] combined the hybrid distance and the Gaussian kernel to construct a novel FRS, and presented the incremental algorithms for feature selection. The efficiency for updating feature sets can be improved, but the variation of multi-attributes was not taken into full consideration. Maji and Garai [28] defined the lower and upper relevance and significance of features for Interval type 2 (IT2) fuzzy approximation spaces, and presented an IT2 FRS-based attribute selection method by integrating the merits of IT2 FRS and the maximal relevance-maximal significance (MRMS) criterion. The effectiveness of the proposed method was shown on several benchmarks and microarray gene expression data sets. Yang *et al.* [29] presented two incremental algorithms with FRS for attribute reduction in terms of one incoming sample and multiple incoming samples, respectively. The relative discernibility relation was incrementally calculated for each condition attribute. The experimental results demonstrated that proposed algorithms could obtain the reduction result with the good classification accuracy. But they were not applied to real-world big data. So

we need to further promote their efficiency on the complex, high-dimensional, and multi-granular big data applications.

B. Limitations and Challenges

In the era of big data, the recent apparent progress of FRS algorithms can be beneficial for the analysis of our confronting big data problems. Meanwhile, it is worth mentioning that, although these algorithms based on FRS are dominant in the classification performance, there is still a lack of deep studies for their applications in current complex big data. The traditional algorithms are suffering from the essential limitations and challenges as follows:

(1) Most of traditional algorithms are more suitable for the medium datasets. If the sample size or the attribute size of the datasets becomes very large, the processing time of attribute reduction will tremendously grow with the increasing of feature dimensions and number of instances. Furthermore, the data dynamism is due to the mechanism that generates related big data changes at different times or different real-world circumstances, which adds new uncertainty for big data analysis. Thus, that inherent interaction relation among different attributes is not fully captured. Although we can incorporate some known information about the desired data partitions into decomposition process, it is not valid for handling dynamic big data tasks. Improving the efficiency of fuzzy rough attribute reduction algorithms in dynamically increasing big data has become a significant research topic, which accelerates the process of finding reduction sets.

(2) The noise problem is one of the main sources of uncertainty in big data applications. When adding noisy or inconsistent data sets that have a lot of boundary objects, most of traditional algorithms usually result in some undesirable feature subsets since their auxiliary space will occupy a large amount of memory, which will be detrimental to the attribute reduction process. Moreover, data objects are normally associated with complex classification scenarios. With the dramatically increasing noise, the speed and volume performances of data generation will deteriorate rapidly. So it is very difficult to generate accurate fuzzy similarity relations for the effective process of fuzzy attribute reduction. The result is often unable to guarantee that the desired reduction set is the optimal attribute set which satisfies the user's need. Consequently, the performance of attribute reduction is often unreliable in most cases. Clearly, the noise problem in big data greatly restricts the practical applications of traditional algorithms.

C. Contributions

In order to address the challenges listed above, we present a new co-evolutionary fuzzy attribute order reduction (CFAOR) algorithm based on complete attribute-value space tree (CAST) to develop the efficient attribute reduction performance for the high-dimensional and uncertain big data. CAST of decision table in the attribute space can adaptively prune and optimize the attribute order tree. The reduced attribute set can satisfy the needs of users with a better convergence speed, which provides the same classification performance with the original attribute set. This CFAOR algorithm is not only suitable for the changing large-scale datasets with interdependent and overlapping attribute variables, but also satisfies the large-scale complex noisy data, which can preserve the

consistency of a given decision table. This CAST provides a new viewpoint to understand and extend the FRS-based fuzzy attribute reduction of big data.

CFAOR is widely compared with state-of-the-art fuzzy attribute reduction methods on publicly-available datasets. The experimental results demonstrate the superiority of CFAOR. CFAOR is applied to identify different tissues surfaces of dynamical changing infant cerebral cortex and it can find more preferred different tissue surfaces from dynamical changing infant cerebrum regions. These encouraging results can achieve the satisfying consistency with those of medical experts. So the main advantages of CFAOR are the high efficiency and robustness for attribute reduction solutions, which makes it particularly suitable for complex big data.

D. Organization

This paper is organized as follows: In Section II, we provide some preliminaries. A CAST model of decision table is constructed in Section III. A new CFAOR algorithm is presented in Section IV. An extensive experimental evaluation is provided in Section V. The application performance in the tissues extraction of dynamical changing infant cerebrum regions is detailed in Section VI. Finally, some conclusions are given in Section VII.

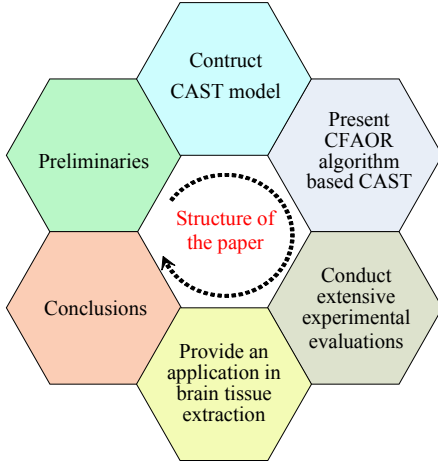


Fig.1. The structure of the paper.

II. PRELIMINARIES

In this section, we introduce some relevant preliminaries related to fuzzy rough set theory and decision-making system with rough entropy threshold.

Definition 1 [10] Let U be a non-empty finite set of samples. Each sample is described by a set of real-valued condition attributes A and a symbolic decision attribute set $D=\{d\}$. The pair $(U, A \cup D)$ and $A \cap D \neq \emptyset$ is called a fuzzy decision table. If the decision attribute d divides U into a family of disjoint subsets $U/D = \{[x]_D : x \in U\}$, $[x]_D = \{y \in U : d(x) = d(y)\}$ is denoted as the decision class to which the sample x belongs.

Definition 2 [8] Generalization of the granule based approximation operators can be obtained by replacing the partition (U/E) by a covering of U . Let I be an index

set, then a collection $\mathbb{C} = \{K_i \subseteq U \mid i \in I\}$ of non-empty subsets of U is called a covering of U if $\bigcup_{i \in I} K_i = U$.

Definition 3 [30][31] A pair of approximation operators is called as dual, if for all $A \subseteq U$, $\overline{apr}(co(A)) = co(\underline{apr}(A))$, where $co(\bullet)$ denotes a covering operation of universe of objects. Equivalence classes can be generalized by neighborhood operators. A neighborhood operator N is a mapping $N : U \rightarrow P(U)$, where $P(U)$ represents the collection of subsets of U . It is assumed that the neighborhood operator is reflexive, i.e., $x \in N(x)$ for all $x \in U$.

Definition 4 [32] 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 $R_a(x, x) = 1$, symmetric $R_a(x, y) = R_a(y, x)$, and sup-min transitive as

$$R_a(x, y) \geq \sup_{z \in U} \min\{R_a(x, z), R_a(z, y)\}, \forall x, y \in U. (1)$$

A subset $B \subseteq A$ can also define a fuzzy equivalence relation, denoted by

$$R_B = \bigcap_{a \in B} R_a, (2)$$

where R_a is a fuzzy equivalence relation.

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 as

$$\underline{R}_B(X)(x) = \inf_{u \in U} \max\{1 - R_B(x, u), X(u)\}, (3)$$

$$\overline{R}_B(X)(x) = \sup_{u \in U} \min\{R_B(x, u), X(u)\}. (4)$$

$\underline{R}_B(X)(x)$ is considered as the degree of x certainly belonging to X , while $\overline{R}_B(X)(x)$ is the degree of x possibly belonging to X . $(\underline{R}_B(X), \overline{R}_B(X))$ is referred to as the fuzzy rough set of X with respect to B .

The essences of lower and upper approximations are demonstrated in Fig. 2.

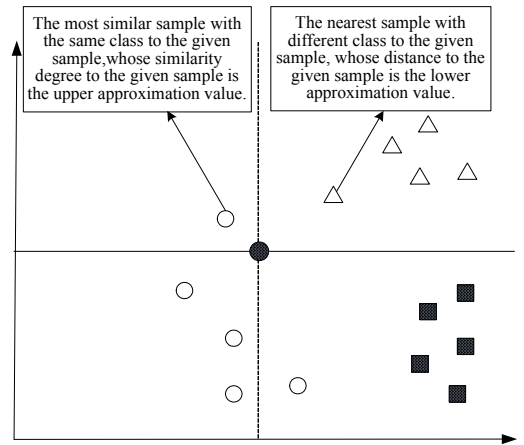


Fig.2. Diagrammatical representation of fuzzy rough approximation operator.

Definition 5 For a fuzzy decision table $(U, A \cup D)$, and $B \subseteq A$, the fuzzy-rough positive region of D with respect to B is defined as

$$Pos_B(D) = \bigcup_{x \in U/D} R_B(x). \quad (5)$$

Definition 6 An attribute subset $P \subseteq A$ is called a reduct of A relative to D , if the following conditions are satisfied:

$$(i) \text{ For } (\forall x \in U), R_P([x]_D)(x) = R_A([x]_D)(x); \quad (6)$$

(ii) For $(\forall a \in P), \exists y \in U$ which satisfies with

$$R_{P-\{a\}}([y]_D)(y) < R_A([y]_D)(y). \quad (7)$$

So a reduct P is a minimum attribute subset of condition attributes that keeps the positive region of D with respect to A . It can discern these sample pairs, of which the corresponding discernibility attribute sets are not empty.

Definition 7 For the fuzzy-rough attribute reduction process, it must be able to find the dependency between various subsets of the original feature set to deal with multiple features. It is necessary to determine the degree of dependency of the decision attribute with respect to $P = \{a, b\}$. In the fuzzy case, since objects may belong to several equivalence classes, the cartesian product of $U/IND(\{a\})$ and $U/IND(\{b\})$ must be considered to determine U/P as follows:

$$U/P = \otimes \{a \in P : U/IND(\{a\})\}. \quad (8)$$

Definition 8 For a fuzzy decision table $(U, A \cup D)$ and the condition attribute set $A = \{a_1, a_2, \dots, a_m\}$, the attribute order satisfying the user's requirements under A is denoted as $S(A) : a_1 \prec a_2 \prec \dots \prec a_k$. Therefore, the optimal reduction model of attribute order is defined as follows:

$$F(x) = \text{Max}(f_s(R)) \quad (9)$$

$$\text{s.t. } R = g(Q), (Q \in A(S), R \in A(R)),$$

where $g : A(S) \rightarrow A(R)$ is any output reduction of attribute sequence with satisfying the needs of users, and R is the reduction of decision table

Definition 9 Supposed a fuzzy decision table $(U, A \cup D)$, $Q \subseteq A$ as a subset of condition attributes, $U/Q = \{X_1, \dots, X_s\}$ and $U/d = \{Y_1, \dots, Y_t\}$. Let the rough entropy threshold be μ ($0.75 \leq \mu \leq 0.95$). The decision rules set with μ is defined as follows:

(i) If a decision rule class $Y_j (1 \leq j \leq t)$ is absolutely rough set with the indiscernible relation Q , then the rule set

$$Q \xrightarrow{\mu} d \text{ is an absolute rough decision rule set with } \mu.$$

(ii) If a decision rule class $Y_j (j > t)$ is relatively rough set with the indiscernible relation Q , then the rule set

$$Q \xrightarrow{\mu} d \text{ is a relatively rough decision rule set with } \mu.$$

III. COMPLETE ATTRIBUTE-VALUE SPACE TREE MODEL

In this section, we present a new optimization model of complete attribute-value space tree structure for fuzzy rough attribute order reduction to find the optimal solution. This tree can adaptively adjust the topological structure of attribute complete tree, and it can successfully finish pruning and

optimizing the attribute order tree for the high-dimensional and uncertain big data. The reduced attribute set can satisfy the needs of users with maintaining a good diversity and a high convergence speed, which provides the same classification performance as the original attribute set.

Definition 10 Basic attributes-value tree denotes that each node of tree is assigned an attribute associated with a basic category, and each branch of node is assigned a value in the range of node attribute's value. The related attributes from the root node to any leaf nodes are satisfied with the given related attribute order.

Given a fuzzy decision table $(U, A \cup D)$, the attributes order $S(B)$ is defined as

$$S(B) : a_1 \prec a_2 \prec \dots \prec a_k, (B \subseteq A, B = \{a_1, a_2, \dots, a_k\}). \quad (10)$$

Attributes-value tree with $S(B)$ is a descending order tree, in which each non-leaf node is given an attribute value a_i of B , each branch is assigned a value in the range of a_i , and each node is associated with a subset of U .

Definition 11 Given a fuzzy decision table $(U, A \cup D)$ and condition attribute subset $B \subseteq A$, $T(a, S(B), U)$ represents the complete attribute-value tree of $S(B)$ in U . $T_D(d_i)$

is used to represent the corresponding leaf nodes a_{+t} in $T(a, S(B), U)$, which can be also denoted as $T_D(a, S(B), U)$.

So the subtree on the root node a_{it} is represented as $T_D(a_{it})$.

Definition 11 Supposed that complete attribute-value space tree (CAST) can be represented as a n -order subtrees $\{T_1, \dots, T_i, \dots, T_n\}$, as outlined in Fig. 3, where $Subpop_i$, Par_i , and $Elit_i$ refer to the i^{th} subpopulation, the i^{th} parent node and the i^{th} elitist node in T_i , respectively,

This CAST can self-adapt the subpopulation sizes in different subtrees and it is employed to capture the interacting attribute order variables by exploiting deep correlation and interdependency among interacting attributes order subsets of big data.

Initially, all co-evolutionary particles are assigned in each node of the original attribute-value space tree, and each inner branch is regarded as a subpopulation with the same number of nodes. As depicted in Fig. 3, it contains two kinds of nodes. One is the ordinary particle, denoted by white hollow dot '○', and the other one is the elitist particle, denoted by the black entity dot '●', which is the best children node in each subtree.

In order to select the best elitist in each subpopulation, particles will be compared by their fitness in attribute-value space tree. In each iteration, each Par_i in $Subpop_i$ is compared to $Elit_i$ in T_i . If $f(Elit_i) < f(Par_i)$, this elitist node of this CAST will be moved up and be exchanged with the parent node. This procedure is continued until all elitists are selected.

Due to excessively pursuing the elitists, CAST regardless of selection directions easily results in the pruning and optimization of attribute order tree into the opposite direction to the sink. Thus, the length of CAST is increased. With the

extension operators, the adjacent elitists will be integrated to reconstruct a unified elitist attribute-value space tree.

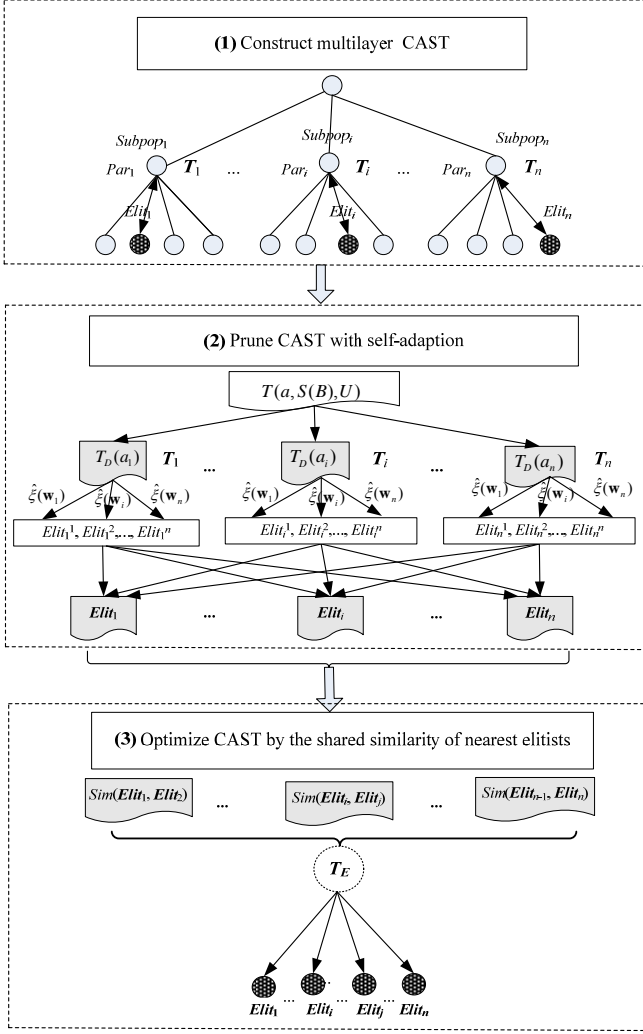


Fig. 3. Construction processes of CAST framework.

The main construction processes of CAST are described as follows:

Algorithm 1: Complete attribute-value space tree (CAST)

1. According to the attribute order of $S(B)$, all nodes in the i^{th} layer ($i \subseteq [1, k]$) are assigned attribute value a_i . The non-leaf nodes in the same level take the same attribute and all leaf nodes are placed in the $(k+1)^{\text{th}}$ layer.
2. Each node in the $(i+1)^{\text{th}}$ layer is associated with an equivalent class as $E_{(i+1)} \in U / \{a_1, a_2, \dots, a_i\}$ ($E_{(i+1)} \neq \emptyset$).
3. If the i^{th} node in the i^{th} layer is associated with the equivalence class as E_i and the child nodes in the $(i+1)^{\text{th}}$ layer is associated with the equivalence class as $E_{(i+1)p}$, the associated equivalent classes of the child nodes are mutually disjoint as $E_{(i+1)} = \bigcup_p E_{(i+1)p}$.
4. $T_D(a_i)$ is a subtree on $T(a, S(B), U)$, and the rules of trimming $T_D(a_i)$ from $T(a, S(B), U)$ are pruned as follows:
 - (i) Calculate $E_{+k} = \bigcup_p E_p$, where E_p is the equivalence classes associated with all leaf nodes in $T_D(a_i)$, and

remove all branches of $T_D(a_i)$. So the node a_i is the leaf node a_{+k} of the associated equivalence class E_{+k} .

- (ii) Replace a_{+k} by the subtree $T_D(d_k)$ according to features of $T(a, S(B), U)$, and continue to prune the corresponding subtree.

5. Adopt the truncated basis to optimize the underlying structure of attribute order tree by

$$\mathbf{x} = \sum_i^n u_i \mathbf{w}_i, \quad (11)$$

where n is the number of variables, \mathbf{w}_i is the selected loading vector and u_i is its corresponding coefficient.

6. Select the ‘best- n -basis’ of attribute order tree in the descending normalized energy score as follows:

$$\hat{\xi}(\mathbf{w}_i) = \mathbf{w}_i^T \mathbf{X} \mathbf{X}^T \mathbf{w}_i, \quad (12)$$

where $\mathbf{X} \in R^{d \times n} = [x_1, x_2, \dots, x_n]$ is the training data matrix of branches of n -order subtrees.

7. Generate a set of reference vector $\Lambda = \{\lambda^1, \lambda^2, \dots, \lambda^n\}$ where

$$\lambda^i = (\lambda_1^i, \lambda_2^i, \dots, \lambda_n^i), \quad (13a)$$

$$\lambda_j^i \in \left\{ \frac{0}{H}, \frac{1}{H}, \dots, \frac{H}{H} \right\}, \sum_{j=1}^n \lambda_j^i = 1, \quad (13b)$$

where H is the number of divisions set along each branch in T_i .

8. Perform the assignment of the subtree neighbourhood $E(i) = \{i_1, i_2, \dots, i_T\}$, $(\lambda^{1T}, \lambda^{2T}, \dots, \lambda^{iT})$ is the closest vectors to λ^i based on the Euclidean distance in T_i .
9. Adopt the Spearman rank correlation coefficient to calculate the similarity between two nearest elitists by

$$\text{Sim}(\text{Elit}_i, \text{Elit}_j) = 1 - \lambda^i \sum_{p=1}^n \frac{(\text{Elit}_i^p - \text{Elit}_j^p)^2}{n(n^2 - 1)}. \quad (14)$$

10. Refine the average shared similarity based on all pairwise similarities by

$$SS = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n (\text{Sim}(\text{Elit}_i, \text{Elit}_j)). \quad (15)$$

11. Reconstruct the complete attribute-value space tree T_E by the elitist sets with the ‘best- n -basis’ of attribute order tree as

$$\text{Elit} = \frac{1}{SS} \{\text{Elit}_1, \text{Elit}_2, \dots, \text{Elit}_n\}, \quad (16)$$

where $\text{Elit}_i = (\text{Elit}_i^1, \text{Elit}_i^2, \dots, \text{Elit}_i^n)$ is the elitist sets of T_i for the ‘best- n -basis’ from the top level basis B_{p-1} as the p projection vectors $(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p)$.

IV. FUZZY ATTRIBUTE ORDER REDUCTION ALGORITHM BASED ON CAST

Since CAST structure of decision table in the attribute space adaptively prunes and optimizes the attribute order tree, the optimal attribute order reduction set can be achieved efficiently. The fuzzy similarity of multi-modality attributes can be fully extracted to satisfy the needs of users with the better convergence speed and classification performance. In this section, we present a co-evolutionary fuzzy attribute order reduction (CFAOR) algorithm for large datasets, especially for

their dynamic and uncertain analysis. And thus, the decision rule sets generate a series of rule chains to develop an efficient cascade attribute order reduction and classification with a rough entropy threshold. Fig. 4 illustrates the framework of attribute order reduction process based on CAST. The main steps of CFAOR are described as follows:

Algorithm 2: Co-evolutionary fuzzy attribute order reduction based on CAST (CFAOR)

Input: The fuzzy decision table $(U, A \cup D)$, $B \subseteq A$, $S(B)$ is the property order of attribute subsets based on fuzzy attribute order $S(A)$, $S(B): a_1 \prec a_2 \prec \dots \prec a_k$, and r_i is the niche neighborhood radius.

Output: Optimal reduction set of fuzzy attribute order RED_{ors} .

1. Initialize the search space of co-evolutionary particles. Create n subpopulations with m particles in each particle subpopulation for optimizing the respective assigned attribute subset.
 2. Divide the sample dataset into two equal parts, one as the training dataset and the rest as the test set. Normalize the attribute order of the training samples and map them into the $[0, 1]$ binary space.
 3. Decompose the large-scale attribute order set into their respective particle subpopulations, and then compute the equivalence classes of their decision tables.
 4. Map the attribute order space $C(A)$ into reduction space $C(R)$, and construct subpopulation with niche neighborhood radius r_i .
 5. Generate complete attribute-value space tree by using **Algorithm 1** with the tree root node a_1 , and $V(a_1) = \emptyset$ in each particle subpopulation $PS_pop(i)$.
 6. Partition the decision system $S = (U, A, V, f)$ into one master-table $S^Q = \langle U, Q, V^Q, f^Q \rangle$ and sub-tables $S^{B_i} = \langle U^{B_i}, B_i \cup \{b_i\}, V^{B_i}, f^{B_i} \rangle, i = 1, 2, \dots, k$.
 7. While (Each joint attribute order $b_i \in Q (i = 2, \dots, k)$) do
 If $(U / IND(Q - \{b_i\})) = U / IND(Q)$,
 then b_i is dispensable and delete it from S^Q and
 $\{Q \leftarrow Q - \{b_i\}, P \leftarrow (P - B_i)\}$
 8. Do
 { (i) Select out the current node a_{im} .
 (ii) If $(a_i(x) \in V(a_{im}))$
 Search the sub-node $a_{(i+1)p}$ of the branch $a_i(x)$, and
 $|a_{(i+1)p}| \leftarrow |a_{im}|$.
 (iii) Else
 { Create a new $a_i(x)$, and set its branch sub-node as
 $a_{(i+1)p}$.
 $V(a_{im}) = V(a_{im}) \cup \{a_i(x)\}$, and $V(a_{(i+1)p}) \neq \emptyset$.
 }
 (iv) $a_m \leftarrow a_{(i+1)p}$.
 } While $(a_i = a_k)$.
 9. Do
 { (i) Reduce $T(a_i, S(B))$ by the defined pruning rules and obtain a attribute-value space tree $T_b(a_i, S(B))$.
 (ii) Choose the max sub-attribute a_k and $R \leftarrow R \cup \{a_k\}$.
 (iii) Push each attribute-value into the stack.
-

(iv) Pop attribute-value behind a_k from the stack in attribute order $S(B)$.

(v) Set a_k as the top stack of attribute order $S(B)$.

(vi) Perform $B \leftarrow \{a_i, a_{i+1}, \dots, a_k\}$ and

$S(B): a_k \prec a_i \prec a_{i+1} \dots \prec a_{k-1}$.

} While $(R = B)$.

10. For (Each reduction set RED_i of $PS_pop(i)$)

{ (i) Move the attribute order $S(C)$.

(ii) Construct the criterion attribute order as

$S_R: a_{i_1} \prec a_{i_2} \prec a_{i_3} \dots \prec a_{i_m}$ and $P \leftarrow \{a_s(x) | a_s \in R\}$.

(iii) $\tilde{P} \leftarrow \{Q | Q \text{ denotes a new searching node by moving } a_s \text{ in different nearest elitists.}\}$

}

11. In the reduction space $C(R)$ of complete attribute-value space tree T_i , optimize the attribute order reduction set as follow:

{ (i) do $\{E(S_{R_i}) \leftarrow F(S_{R_i}) - \tilde{P}$.

$Q \in E(S_{R_i})$ and $R \leftarrow g(Q)$.

If $(f_s(R) \geq f_s(S_{R_i}))$

$\{R \leftarrow R_i, E(S_{R_i}) \leftarrow E(S_{R_i}) - Q\}$

} While $(E(S_{R_i}) = \emptyset)$.

(ii) $\forall r_j \in R, R \leftarrow (R - S_{R_j})$, and compute $IND_{S_R \cup r_j}^k$.

(iii) $S_R \leftarrow S_R \cup S_{R_j}$, where

$$|IND_{S_R \cup r_m}^k| = \min(|IND_{S_R \cup r_j}^k|).$$

}

12. Compose a mid-table S^{M_i} with both the sub-table S^{B_i} and the master-table S^Q .

13. While (Each attribute $a \in B_i$) do

If $(U / IND(M_i - \{a\}) \neq U / IND(M_i))$, then a is indispensable and perform

$\{B_i \leftarrow (B_i - \{a\}), Red \leftarrow Red \cup B_i, \text{ and } P = P / Red\}$

14. Evaluate whether the accuracy of fuzzy attribute order reduction is satisfied with the rough entropy threshold μ or not.

If satisfied, output the optimal reduction set RED_{ors} of fuzzy attribute order.

Otherwise, return to **Step 9** to continue to perform fuzzy attribute order reduction.

V. EXPERIMENTAL RESULTS

In this section, we conduct some experiments to validate the effectiveness and robustness of the proposed CFAOR algorithm. We start with the description of experimental setting in Section A. The performance comparisons with time and space between CFAOR and some representative fuzzy attribute reduction algorithms are conducted in Section B. In Section C, we carry out the accuracy comparisons of fuzzy attribute reduction and classification.

A. Experimental setup

In order that the comparisons of fuzzy attribute reduction performance are more sufficient, we select ten real-world

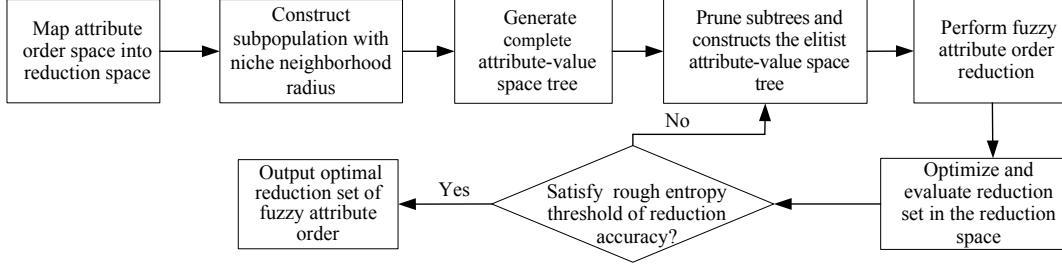


Fig. 4. Framework of fuzzy attribute order reduction process.

datasets from the UCI Machine Learning Repository and NIPS 2003 feature selection challenge datasets [33]. Those datasets, as described in Table I, include five low-dimensional datasets with adding 15% Gaussian noise ratio and five very high-dimensional datasets which come from various real-world domains, such as text categorization (Dexter), radar (Ionosphere), drug science (Dorothea), medical science (Spect and Spectf), and biomedical science (Ovarian-cancer and Lung-cancer). We add 15% and 25% noise ratios to five low-dimensional and five high-dimensional datasets, respectively. All algorithms are implemented in Visual C# 2013, and all experiments are run on a virtual machine with 12 CPUs and 256 GB memory at the University of Technology Sydney (UTS) High Performance Computing Linux Cluster with 8 nodes.

CFAOR is compared to three state-of-the-art algorithms for feature selection. To make the comparisons as fair as possible, the parameters of the compared algorithms are set to be the same as suggested in their respective references. The parameters of CFAOR are initialized as follows: The initial number of co-evolutionary particles is 2000, the number of particle subpopulation is 50, the number of particles in each particle subpopulation is 400, the number of iterations is 1000, and the number of iterations within each subpopulation is 500. The rough entropy threshold is satisfied with $\mu = 92\%$.

TABLE I
SUMMARY OF 10 REAL-WORLD DATASETS

Dataset	#Feature	#Sample	Dataset	#Feature	#Sample
Spect	22	267	Madelon	500	2,000
Mushroom	22	8,124	Dexter	20,000	300
Ionosphere	34	351	Dorothea	100,000	800
Spectf	44	267	Ovarian-cancer	2,190	216
Infant	86	5,339	Lung-cancer	12,533	181

B. Performance comparison metrics with time and space for different algorithms

Table II indicates the performance comparison metrics with Time and Space for different algorithms, where Time is the running time (in seconds), and Space is the space consumption (in Megabytes). Additionally, the bold value means it is the best result among different fuzzy attribute reduction algorithms.

As shown in Table II, it is clear that the best solutions obtained by CFAOR are better than those by NFRS [22], UFRFS [26] and RDRAR[29], although each algorithm comes

with a substantially reduced computing running time and space load consumption. The bigger dataset is the more significant computational savings for CFAOR. Meanwhile, it is remarkable that in our experiments, the big space consumption also leads to the memory overflow for NNFS in Madelon, Dorothea and Lung-cancer datasets, and also for UFRFS in Madelon, Dexter and Dorothea datasets. But our proposed CFAOR algorithm can cope with this situation well. As it can be seen in Table II, CFAOR clearly outperforms NFRS, UFRFS and RDRAR in terms of Time and Space performance metrics for most datasets. As for the Spectf, Infant and Dorothea datasets, the significant improvements are brought by CFAOR. For example, CFAOR spends the 35.20% less space consumption (Space) than NFRS on the Spectf dataset, and 52.12% on the Infant dataset. These facts indicate that CFAOR can find consistently much faster and consuming less memory, compared with other algorithms.

From Table II, one can clearly see that, in term of the time results, CFAOR outperforms all compared algorithms across all instances. Furthermore, both the time and space results obtained by CFAOR are better than NNFS on all instances. The results reveal that the fuzzy attribute order reduction based on complete attribute-value space tree contributes to CFAOR performance, which has an effect on the ability in producing high quality results across all testing instances.

C. Accuracy comparisons fuzzy attribute reduction and classification

In this section, we carry out comparisons of fuzzy attribute reduction and classification accuracy of CFAOR and five compared algorithms: NFRS [22], UFRFS [26], IT2-FR [28] FR-MRMS [28], and RDRAR [29]. Using C4.5 [34] as classifier, we employ a stratified 10×10-fold cross-validation (10-FCV). The original dataset is divided into 10 subsets of instances, where one subset is retained as the testing data, and the remaining subsets are used for training data.

Table III depicts the comparisons of the prediction accuracy and standard deviation of all algorithms. We can see that CFAOR is obviously better than the other five techniques. NFRS, RDRAR and IT2-FR cannot deal with high-dimensional datasets, and both RDRAR and NFRS show the worst performance among all algorithms with failing to offer results for most datasets. Meanwhile, it is also noticed that all algorithms do not achieve high prediction accuracy because Madelon dataset contains many noisy features without predictive power.

TABLE II
COMPARISONS OF PERFORMANCE METRICS WITH TIME AND SPACE FOR DIFFERENT

Datasets	NFRS		UFRFS		RDRAR		CFAOR	
	Time (s)	Space (M)	Time (s)	Space (M)	Time (s)	Space (M)	Time (s)	Space (M)
1. Spect	2.210	1.980	3.107	2.154	2.631	1.489	1.256	1.031
2. Mushroom	2.328	2.843	2.194	2.186	2.006	1.098	1.190	1.115
3. Ionophere	3.284	2.356	4.169	3.817	3.087	2.124	1.689	1.225
4. Spectf	4.183	2.092	4.927	1.678	3.288	2.134	1.781	1.787
5. Infant	13.165	8.560	14.287	12.978	12.098	8.327	8.189	6.237
6. Madelon	<i>Memory Overflow</i>		<i>Memory Overflow</i>		28.110	21.092	20.324	17.390
7. Dexter	18.604	19.217	<i>Memory Overflow</i>		19.176	17.693	16.325	13.098
8. Dorothea	<i>Memory Overflow</i>		<i>Memory Overflow</i>		45.672	25.138	31.768	19.675
9. Ovarian-cancer	16.063	16.117	16.170	14.187	15.675	13.430	13.290	10.190
10. Lung-cancer	26.137	20.280	42.217	25.169	40.542	22.007	34.789	18.767

TABLE III
Classification accuracy of CFAOR and its rivals using C4.5 classifier. (%) (Test \pm Std /%)

Datasets	UFRFS	RDRAR	NFRS	IT2-FR	FR-MRMS	CFAOR
Spect	83.43 \pm 0.29	84.16 \pm 0.57	87.98 \pm 0.23	86.89 \pm 0.32	88.21 \pm 0.25	92.17 \pm 0.08 δ
Mushroom	97.75 \pm 0.81	96.22 \pm 0.43	96.39 \pm 0.31	97.35 \pm 0.24	96.09 \pm 0.18	100 \pm 0
Ionophere	80.29 \pm 0.63	83.89 \pm 0.21	86.17 \pm 0.28	84.18 \pm 0.42	85.20 \pm 0.11	91.79 \pm 0.05 δ
Spectf	81.12 \pm 0.39	81.12 \pm 0.23	85.27 \pm 0.23	84.13 \pm 0.20	84.89 \pm 0.13	89.89 \pm 0.44
Infant	77.12 \pm 0.71	76.11 \pm 0.21	79.90 \pm 0.39	82.00 \pm 0.15	83.98 \pm 0.21	87.48 \pm 0.49 δ
Madelon	52.25 \pm 0.39	67.11 \pm 0.93	70.21 \pm 0.67	68.13 \pm 0.24	70.18 \pm 0.31	74.93 \pm 0.08
Dexter	76.18 \pm 0.52	70.09 \pm 0.74	83.68 \pm 0.26	78.37 \pm 0.53	83.43 \pm 0.24	87.92 \pm 0.09 δ
Dorothea	85.11 \pm 0.18	—	—	86.17 \pm 0.45	87.21 \pm 0.32	95.56 \pm 0.52 δ
Ovarian-cancer	96.38 \pm 0.62	—	—	—	94.98 \pm 0.35	100 \pm 0 δ
Lung-cancer	93.27 \pm 0.19	—	—	—	95.34 \pm 0.43	97.68 \pm 0.65

Notes: ‘—’ denote that the algorithm cannot deal with high dimensional dataset due to expensive computations. ‘ δ ’ denote that the significantly best mean value is highlighted in boldface with gray background.

As shown in Table III, CFAOR with C4.5 classifier has 92.17%, 91.79%, 95.56% and 100% significant average classification accuracies on Spectf, Ionophere, Dorothea and Ovarian-cancer datasets, respectively, whose corresponding results are identified as the symbol ‘ δ ’. If related algorithms cannot deal with high dimensional datasets due to expensive computations, their results are identified as the symbol ‘—’.

It is obvious that CFAOR is consistently much better than its competitors on almost datasets. Most compared algorithms simply select a few of features from the correlated features set, whereas CFAOR considers both strongly relevant features and their corresponding correlated features simultaneously, which will turn out to be beneficial to reduce the classification error. The main reasons for this are as follows: We construct a complete attribute-value space tree structure of decision table in the attribute space to adaptively prune and optimize the

attribute order tree. It can perform attribute order reduction in a much shorter time, and the reduced attribute set can satisfy the needs of users guaranteeing a better convergence speed. Therefore, CFAOR can well avoid some recalculations by directly exploiting the previous results from the current accumulated samples. Of course, it is also necessary to point out that in few special cases, the performance of CFAOR is slightly worse than those compared algorithms. The experimental results obviously demonstrate that the classification system by employing CFAOR as the fuzzy attribute reduction algorithm leads to the appealing performance of classification accuracy. CFAOR provides an effective approach to obtain the optimal result of fuzzy attribute reduction, which significantly enhances the classification accuracy with a reinforcing noise tolerance.

VI. APPLICATION TO TISSUES EXTRACTION OF DYNAMICAL CHANGING INFANT CEREBRAL CORTEX

Various tissues extraction of infant cerebral cortex is a very useful technique for the assessment of different regional matters in brain neurodegenerative diseases. The classification of the neonatal brain MRI is a critical step towards the infant brain development. During the postnatal human brain development, the brain tissues often undergo a wide range of development. Because there are some distinct differences of brain tissue structures between the infant and adult, state-of-the-art methods for adult brain classification are not applicable to the infant brain, which poses additional challenges for the study of infant cerebral cortex [35]-[37].

In this selection, we employ CFAOR to accelerate the tissue extraction of dynamical changing infant cerebral cortex with better accuracy and efficiency. We obtain infant brain datasets from the Internet Brain Segmentation Repository (IBSR) [38]. We apply different algorithms to determine the brain volumes of these subjects and then employed the manually segmented brain areas to evaluate the extraction accuracy. CFAOR is compared with five popular methods (BET [39], BSE [40], FreeSurfer [41], LongSeg [42], BEaST [43], and LPG- PCA [44]). The default parameters of these methods are used with their respective references.

A. Extraction results of infant brain tissues from the inner, outer and central cortical surfaces

To reveal the segmentation accuracy of cortical surfaces for infant brain 3D-MRI, we estimate of segmentation results from the inner, outer and central cortical surfaces. We label the matters inside the inner neonatal cortical surface as WM, the matters between the inner and outer cortical surfaces as GM, and the matters between the skull and outer cortical surfaces as CSF. For ease of viewing, only part of the segmentation results is shown in Fig. 5. It can be seen that there are some visible isolated holes caused by BET and BSE, but CFAOR can smooth out most of imperfections with disposing of them as parts of the cortical surfaces, and it shows the essential characteristic, which is consistent with the real inner and outer cortical surfaces.

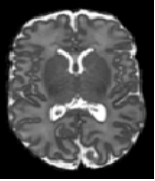

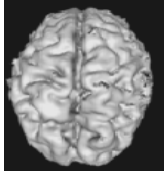

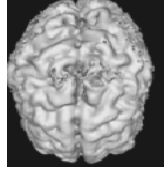


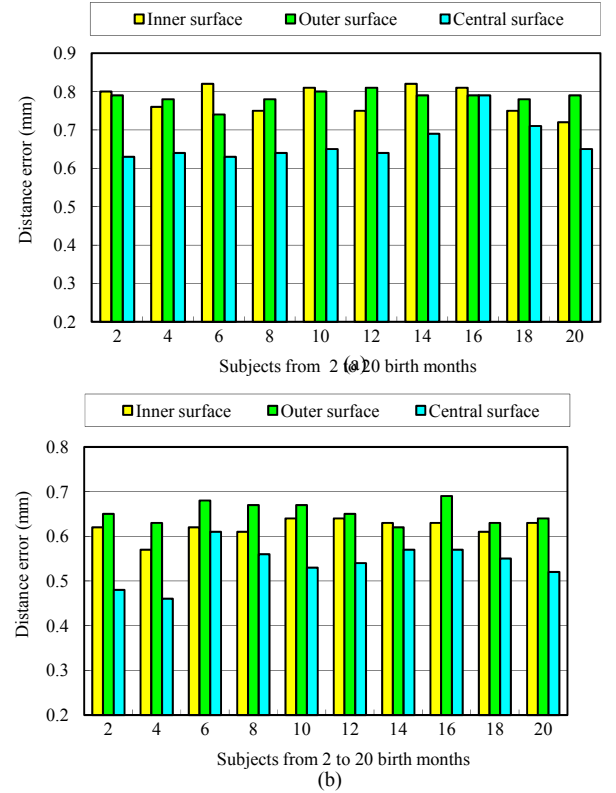
 <p>Infant brain 3D-MRI slice by added 15% noise ratio and 20% INU level</p>			BET
			BSE
			CFAOR
	Inner cortical surfaces	Outer cortical surfaces	Methods

Fig. 5. 3D rendering cortical reconstruction comparisons of inner and outer cortical surfaces.

In the following, the average surface distance error of different cortical surfaces is further investigated. Since there are large developmental changes in the developing infant brain matters, we perform the subjects at different birth months to validate the robustness of different algorithms. We select 10 subjects from 2 to 20 birth months from 0.10% 3D MR-18 dataset size. The average distance errors of 10 subjects are illustrated in Fig. 6, where those of the inner, outer and central cortical surfaces are around 0.779mm/0.785mm/0.667mm (BET), 0.701mm/0.728mm/0.592mm (BSE), 0.620mm/0.658mm/0.539mm(CFAOR), respectively. CFAOR achieves overall 9.8%-25.6% improvement. Especially, CFAOR shows a better mean distance error of central cortical surface (0.539 mm), which reflects its significantly higher sensitivity and accuracy. Although, BSE also achieves good results, but it needs to speed more computational cost to obtain those compromised results. Thus, CFAOR can use crowns of gyri to properly simulate growth with simple image warping, and generate a more accuracy results.

From Fig. 6, it is obvious that CFAOR achieves the highest accuracy, indicating its superiority in better characterizing the structural longitudinal surfaces of infant cerebral cortex. Therefore, it can significantly improve the mean classification performance, but two compared methods are under-classification for cerebral regions because some deformable surfaces are located inner cerebral surface regions.



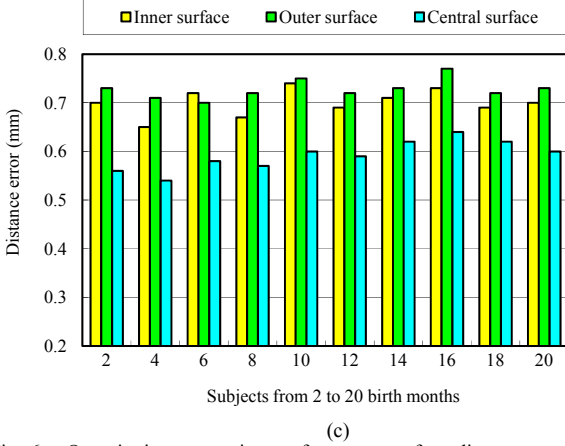


Fig. 6. Quantitative comparisons of average surface distance errors of the inner, outer and central cortical surfaces for 10 infant brain 3D MR-18 subjects. (a) BET, (b) BSE, and (c) CFAOR.

B. Quantification comparisons of multi-criteria performance

High resolution and high contrast infant MRI allow for the tissue extraction of dynamical changing human brain cerebral cortex, which can lead to successful and accurate segmentation. In the following experiments, multi-criteria for brain tissue extraction assessment, such as the Jaccard similarity coefficient (JSC), the specificity and sensitivity coefficients, and the missed extraction probabilities, are employed to measure the comprehensive performance.

JSC is widely adopted to evaluate the similarity between the extracted brain region Y and the corresponding ground truth X as

$$JSC(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}, \quad (17)$$

where $|\cdot|$ denotes the cardinality value, and the value of JSC is within $[0, 1]$.

The high sensitivity is regarded as the high recognizing percentage for cerebral cortex, denoted as S_e . The high specificity is regarded as the high rejecting percentage for non-brain cerebral cortex, denoted as S_p . Brain tissue extraction is a compromise between S_e and S_p . Their coefficients are defined as

$$S_e = \frac{TP}{TP + FN}, \quad S_p = \frac{TN}{TN + FP}, \quad (18)$$

where TP is the true positive rate which is the number of voxels correctly classified as brain cerebral cortexes, FP is the false positive rate which is the number of voxels incorrectly classified as brain cerebral cortexes, TN is the true negative rate which is the number of voxels correctly classified as non-brain tissues, and FN is the false negative rate which is the number of voxels incorrectly classified as non-brain cerebral cortexes.

Furthermore, we adopt two probabilities of missed extraction for both brain tissues p_m and false alarm p_f to measure the extraction risk, which are calculated as

$$p_m = \frac{|X - Y|}{|X \cup Y|}, \quad p_f = \frac{|Z - X|}{|X \cup Y|}, \quad (19)$$

where Z is the extracted brain region with false alarm.

Table IV shows the experimental results of CFAOR and compared popular methods on the IBSR datasets. Jointly considering S_e and S_p , the accuracy of BET, BSE and LongSeg are moderate. However, CFAOR is able to correctly detect the dynamical changing infant cerebral cortex and became remarkable improvement in terms of sensitivity and specificity. CFAOR often performed better than compared popular methods with multi-criteria performance, which indicates its better overlapping of the extracted brain regions with the ground truths. We also notice that LPG-PCA can achieve the better sensitivity in detecting almost brain tissues at the expense of the relatively low specificity.

Meanwhile, it is noticeable from Table IV that the computational time of CFAOR is obviously lower than BSE, BEaST, and FreeSurfer for all tested instances. CFAOR produces high quality solutions with lower computational times, followed by LPG-PCA, BET and LongSeg.

Through these restricted analyses, we can conclude that the extraction of brain tissues yielded by CFAOR further facilitates the correction of intensity non-uniformity for dynamical changing infant cerebral cortex.

C. Comparison of the Dice similarity coefficient of expert consensus extraction

To further ensure the reliability of CFAOR, we compare the Dice coefficient of our extraction and the expert consensus extraction between each pair of expert manual extraction. We quantitatively examine the overlap level between our extraction and manual expert extraction. Dice similarity coefficient is defined as

$$\text{Dice}(A, B) = \frac{2|A \cap B|}{|A| + |B|} \times 100\%, \quad (20)$$

where A and B are the voxel sets of two different extractions of the same tissue, respectively.

We obtain average Dice values of ten subjects in Table V, which show the differences between our extraction and the expert consensus and the differences of different experts. It also indicates that the committed errors are in the same magnitude range with the inter-experts variability. CFAOR can achieve the highest Dice similarity coefficient as expert consensus extraction and boost much better consistent labeling boundaries for large-scale dynamical changing infant cerebral cortex.

VII. CONCLUSIONS

A. Discussion of experimental results

As illustrated in above experimental results, it is easy to draw the conclusion that CFAOR outperforms its competitors on most of the used complex datasets and achieves the higher computational efficiency and classification accuracy. CFAOR is effective and efficient. Furthermore, it can be well applied to the tissues extraction of dynamical changing infant cerebral cortex and achieves satisfying results. The reported results illustrate that its predication is highly correlated with the human learning and evaluation. CFAOR has very low-time

TABLE IV
Performance comparison for brain extraction algorithms using the IBSR dataset.

Methods	<div>$JSC$$S_e$$S_p$$P_m$$P_f$</div>					Time (s)
	(Test \pm Std /%)					
BET	82.92 \pm 0.39	93.46 \pm 0.58	94.23 \pm 0.64	4.79 \pm 2.37	10.21 \pm 1.08	56.23
LongSeg	82.18 \pm 0.51	94.46 \pm 0.76	94.49 \pm 0.29	5.83 \pm 2.20	21.37 \pm 1.52	61.89
BSE	85.24 \pm 0.875	93.23 \pm 0.63	94.11 \pm 0.66	6.82 \pm 0.88	19.43 \pm 1.42	78.23
BEaST	87.74 \pm 0.15	95.67 \pm 0.87	95.14 \pm 0.29	4.47 \pm 0.59	16.22 \pm 0.32	123.09
FreeSurfer	85.12 \pm 0.69	94.11 \pm 0.37	96.19 \pm 0.84	5.28 \pm 1.06	11.32 \pm 1.21	218.28
LPG-PCA	89.34 \pm 0.32	95.51 \pm 0.71	97.22 \pm 0.43	3.32 \pm 1.19	9.25 \pm 0.35	50.28
CFAOR	92.32 \pm 0.25	97.32 \pm 0.43	99.12 \pm 0.31	2.76 \pm 0.42	7.89 \pm 0.45	32.10

TABLE V
Comparison of the Dice similarity coefficient between our extraction and the expert consensus extraction.

Subjects	Tissue automatic vs. Consensus	Expert 1 vs. Expert 2	Expert 2 vs. Expert 3	Expert 3 vs. Expert 4	Expert 4 vs. Expert 5	Expert 3 vs. Expert 5	Average Dice of five pairs of experts
GM	0.93	0.91	0.93	0.92	0.91	0.91	0.92
WM	0.94	0.89	0.92	0.95	0.93	0.93	0.93
SGM	0.90	0.92	0.87	0.85	0.89	0.89	0.88
MWM	0.84	0.83	0.79	0.83	0.82	0.82	0.82
CSF	0.85	0.85	0.84	0.84	0.85	0.85	0.84

complexity and high accuracy with high quality solutions, compared to the state-of-the-art methods.

In contrast, other compared representative fuzzy attribute reduction algorithms as NFRS [22], UFRFS [26], IT2-FR [28], FR-MRMS [28] and RDRAR [29], load features into memory at one time, which costs are heavy in both memory and time, so that the time cost for these algorithms fluctuate obviously because their computational complexity and their classification performance dynamically decrease as time increases.

CFAOR is robust to the large-scale sample size dataset by added the noisy percentage, which means it is more stable than the compared representative algorithms. CFAOR only requires a relatively small time to achieve remarkable classification performance and it is more suitable to be used on noisy attributes than other representative algorithms. Moreover, CFAOR has better classification performance on five high-dimensional datasets with large noisy percentage attributes.

The traditional fuzzy attribute reduction algorithms are unreliable for dealing with the dynamically changing massive datasets with ever-greater amounts and complex fuzzy structures. However, CFAOR is more advantageous than traditional methods to measure accuracy of attribute reduction and classifications in dynamically changing uncertain big data. The major reasons are as follows: the use of fuzzy attribute order reduction model based on CAST structure can explore the structure of the dataset in the attribute dimension, which will reduce the effect of the noisy attributes with good speed. Therefore, it greatly improves the search efficiency in finding the

optimal solution. More importantly, CAST can adaptively adjust the topological structure of attribute complete tree, and it can successfully finish pruning and optimizing the attribute order tree for the high-dimensional and uncertain big data. The reduced attribute set can satisfy the needs of users with maintaining a good diversity and a high convergence speed. Nevertheless, the results obtained by compared representative algorithms are not satisfactory. They are often not suitable for the attribute reduction classification task in the above-mentioned complex real-world datasets described in Introduction.

Therefore, we observe that CFAOR has a great superiority in terms of computational time and accuracy of fuzzy attribute reduction, especially for the high-dimensional and uncertain large-scale datasets to be processed. The classification superiority of CFAOR has been clearly proved when big datasets are added the noisy percentages, while its accuracy values maintain stable, which are depicted in Tables II and III. In summary, we can conclude CFAOR is a better choice to balance the computational cost and accuracy of fuzzy attribute reduction, compared with representative algorithms.

B. Closure

In this paper, we have presented a new co-evolutionary fuzzy attribute order reduction algorithm, CFAOR. This algorithm can deal with lots of uncertain variables of fuzzy attribute sets in big data so that we induce their complexity and non-separability. A complete attribute-value space tree model of decision table is constructed in the attribute space to adaptively prune and

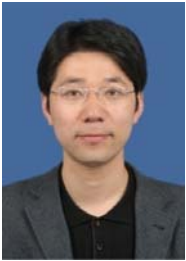
optimize the attribute order tree, with providing the same classification performance as the original attribute set. The experimental results have demonstrated that the proposed CFAOR algorithm can carry out attribute order reduction and classification more accurately. It is able to effectively deal with a variety of forms and distributions of big datasets. CFAOR turns out to be efficient and robust on large-scale attribute reduction classification tasks. Furthermore, it is applied into automatic tissues extraction of human cerebral cortex of infant MRIs, and our results have indicated that CFAOR can be used for reliable tissues extraction in the lower resolution infant cerebral cortex. Hence, the proposed CFAOR algorithm is useful for evaluating neuro-protective clinical trials in infant brain.

Despite of these promising results of tissues extraction of large-scale dynamical changing infant cerebral cortex, the visual comparison of automatic extraction results yielded by CFAOR shows a bit of regions of MWM buried in the gyri that cannot be detected. The main reason is that the partial volume effect makes it difficult for the MWM region to enter narrow channels in the cortical gray matter [45][46]. It limits the CFAOR's application in real-world large-scale infant cerebral cortexes. In the future, we plan to enable the straightforward use of the tissues extraction for an accurate reconstruction of the gradual myelination process, which should allow for a higher improvement in the complex infant cerebral resolution.

REFERENCES

- [1] V. Marx, "Biology: The big challenges of big data," *Nature*, vol. 498, no. 7453, pp. 255–260, Jun. 2013.
- [2] B. Wenjie, M. Cai, M. Liu, and G. Li, "A big data clustering algorithm for mitigating the risk of customer churn," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 1270–1281, Jun. 2016.
- [3] L. Zhihan, J. Chirivella, and P. Gagliardo, "Bigdata oriented multimedia mobile health applications," *J. Med. Syst.*, vol. 40, no. 5, pp. 1–10, May 2016.
- [4] P. Norvig, D. A. Relman, and D. B. Goldstein, et al., "2020 visions," *Nature*, vol. 463, no. 7, pp. 26–32, 2010.
- [5] S. Madden, "From databases to big data," *IEEE Internet Comput.*, vol. 16, no. 3, pp. 4–6, May/Jun. 2012.
- [6] H. V. Jagadish, J. Gehrke, and A. Labrinidis, et al., "Big data and its technical challenges," *Commun. ACM*, vol. 57, no. 7, pp. 86–94, Jul. 2014.
- [7] J. Wu, Z. Wu, and J. Cao, et al., "Fuzzy consensus clustering with applications on big data," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1430–1445, Dec. 2017.
- [8] W. Pedrycz, *Granular Computing: Analysis and Design of Intelligent Systems*, CRC Press/Francis Taylor, Boca Raton, 2013.
- [9] S. K. Pal, B. U. Shankar, and P. Mitra, "Granular computing, rough entropy and object extraction," *Pattern Recogn. Lett.*, vol. 26, no. 16, pp. 2509–2517, Dec. 2005.
- [10] L. A. Zadeh, "Fuzzy sets," *Inf. Contr.*, vol. 8, no. 3, pp. 338–353, Jun. 1965.
- [11] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, Oct. 1982.
- [12] S. Greco, B. Matarazzo, and R. Slowinski, "Rough sets theory for multicriteria decision analysis," *Eur. J. Oper. Res.*, vol. 129, no. 1, pp. 1–47, Feb. 2001.
- [13] W. Ding, and J. Wang, "A novel approach to minimum attribute reduction based on quantum-inspired self-adaptive cooperative co-evolution," *Knowl. Based Syst.*, vol. 50, pp. 1–13, Sep. 2013.
- [14] H. M. Chen, T. R. Li, and C. Luo, et al., "A rough-set-based method for updating decision rules on attribute values' coarsening and refining," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 12, pp. 2886–2899, Dec. 2014.
- [15] J. Wang, J. Peng, and O. Liu, "A classification approach for less popular webpages based on latent semantic analysis and rough set model," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 642–648, Jan. 2015.
- [16] Y. Yao, Y. She, "Rough set models in multigranulation spaces," *Inf. Sci.*, vol. 327, pp. 40–56, Jan. 2016.
- [17] A. M. Radzikowska and E. E. Kerre, "A comparative study of fuzzy rough sets," *Fuzzy Sets Syst.*, vol. 126, no. 2, pp. 137–155, Mar. 2002.
- [18] Q. H. Hu, D. R. Yu, W. Pedrycz, and D. G. Chen, "Kernelized fuzzy rough sets and their applications," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 11, pp. 1649–1667, Nov. 2011.
- [19] R. Wang, D. G. Chen, and S. Kwong, "Fuzzy-rough-set-based active learning," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 6, pp. 1699–1704, Dec. 2014.
- [20] Y. H. Qian, Q. Wang, and H. H. Cheng, et al., "Fuzzy-rough feature selection accelerator," *Fuzzy Sets Syst.*, vol. 258, pp. 61–78, Jan. 2015.
- [21] D. S. Yeung, D. G. Chen, and E. C. Tsang, et al., "On the generalization of fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 3, pp. 343–361, Jun. 2005.
- [22] C. Cornelis, J. Medina, and N. Verbiest, "Multi-adjoint fuzzy rough sets: definition, properties and attribute selection," *Int. J. Approx. Reason.*, vol. 55, pp. 412–426, Jan. 2014.
- [23] S. Zhao, E. C. C. Tsang, D. Chen, and X. Z. Wang, "Building a rule-based classifier-A fuzzy-rough set approach," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 5, pp. 624–638, May. 2010.
- [24] R. Jensen and C. Cornelis, "Fuzzy-rough nearest neighbour classification and prediction," *Theoretical Computer Science*, vol. 412, no. 42, pp. 5871–5884, Sep. 2011.
- [25] Q. H. Hu, L. Zhang, S. An, D. Zhang, and D. R. Yu, "On robust fuzzy rough set models," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 4, pp. 636–651, Aug. 2012.
- [26] N. M. Parthalain, and R. Jensen, "Unsupervised fuzzy-rough set-based dimensionality reduction," *Inf. Sci.*, vol. 229, pp. 106–121, Apr. 2013.
- [27] A. P. Zeng, T. R. Li, and D. Liu, et al., "A fuzzy rough set approach for incremental feature selection on hybrid information systems," *Fuzzy Sets Syst.*, vol. 258, pp. 39–60, Jan. 2015.
- [28] P. Maji and P. Garai, "IT2 fuzzy-rough sets and max relevance-max significance criterion for attribute selection," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1657–1668, Aug. 2015.
- [29] Y. Yang, D. Chen, and H. Wang, et al., "Fuzzy rough set based incremental attribute reduction from dynamic data with sample arriving," *Fuzzy Sets Syst.*, vol. 312, pp. 66–86, Apr. 2017.
- [30] Z. Bonikowski, E. Bryniarski, and U. W. Skardowska, "Extensions and intentions in the rough set theory," *Inf. Sci.*, vol. 107, nos. 1–4, pp. 149–167, Jun. 1998.
- [31] T. Yang and Q. Li, "Reduction about approximation spaces of covering generalized rough sets," *Int. J. Approx. Reason.*, vol. 51, no. 3, pp. 335–345, Feb. 2010.
- [32] D. Yeung, D. G. Chen, and E. Tsang, et al., "On the generalization of fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 3, pp. 343–361, Jun. 2005.
- [33] M. Lichman, UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences, <http://archive.ics.uci.edu/ml>, 2013.
- [34] R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann, 1993.
- [35] J. E. Iglesias, C. Y. Liu, P. M. Thompson, and Z. Tu, "Robust brain extraction across datasets and comparison with publicly available methods," *IEEE Trans. on Medical Imaging*, vol. 30, no. 9, pp. 1617–1634, Sep. 2011.
- [36] C. N. Devi, A. Chandrasekharan, V. K. Sundaraman, and Z. C. Alex, "Neonatal brain MRI segmentation: A review," *Computers in Biology and Medicine*, vol. 64, pp. 163–178, Sep. 2015.
- [37] L. Gui, R. Lisowski, and T. Faundez, et al., "Morphology-driven automatic segmentation of MR images of the neonatal brain," *Medical Image Analysis*, vol. 16, no. 8, pp. 1565–1579, Dec. 2012.
- [38] The Center for Morphometric Analysis at Massachusetts General Hospital, <http://www.cma.mgh.harvard.edu/ibsr>.
- [39] S. M. Smith, "Fast robust automated brain extraction," *Hum. Brain Mapp.*, vol. 17, no. 3, pp. 143–155, Nov. 2002.
- [40] D. W. Shattuck, and R. M. Leahy, "Automated graph-based analysis and correction of cortical volume topology," *IEEE Trans. Med. Imaging*, vol. 20, no. 11, pp. 1167–1177, Nov. 2001.
- [41] "FreeSurfer," online at <http://surfer.nmr.mgh.harvard.edu>.
- [42] F. Shi, Y. Fan, S. Tang, and J. H. Gilmore, et al., "Neonatal brain image segmentation in longitudinal MRI studies," *NeuroImage*, vol. 49, no. 1, pp. 391–400, Jan. 2010.
- [43] S. F. Eskildsen, P. Coupé, V. Fonov, and J. V. Manjon, "BEaST: brain extraction based on nonlocal segmentation technique," *NeuroImage*, vol. 59, no. 3, pp. 2362–2372, Feb. 2012.
- [44] L. Zhang, W. Dong, and D. Zhang, et al., "Two-stage image denoising by principal component analysis with local pixel," *Pattern Recogn.*, vol.

- 43, no.4, pp.1531-1549, Apr. 2010.
- [45] P. Anbeek, I. Išgum, and B. J. M. van Kooij, et al., "Automatic segmentation of eight tissue classes in neonatal brain MRI," *PLOS one*, vol. 8, no. 12, e81895, Dec. 2013
- [46] T. Ertekin, N. Acer, and S. Icer, et al., "Estimation of the total brain volume using semi-automatic segmentation and stereology of the newborns' brain MRI," *NeuroQuantology*, vol. 11, no. 2, pp. 181-188, Jun. 2013.



Weiping Ding (M'16) received the Ph.D. degree in Computation Application, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2013. He was a visiting researcher at University of Lethbridge, Alberta, Canada, in 2011. From 2014 to 2015, He is a Postdoctoral Researcher at the Brain Research Center, National Chia Tung University (NCTU), Hsinchu, Taiwan. In 2016, He was a Visiting Scholar at National University of Singapore (NUS), Singapore. From 2017 to 2018, he was a Visiting Professor at University of Technology Sydney (UTS), Australia. He is a member of IEEE ACM, and Senior CCF. He has published over 50 papers in flagship journals and conference proceedings as the first author. To date, he has held 10 approved invention patents in total over 18 issued patents. Dr. Ding is currently an Associate Professor, and main research directions involve data mining, machine learning and granular computing. Dr. Ding was a recipient of Computer Education Excellent Paper Award (First-Prize) from the National Computer Education Committee of China, in 2009. He was an Excellent-Young Teacher (Qing Lan Project) of Jiangsu Province in 2014, and a High-Level Talent (Six Talent Peak) of Jiangsu Province in 2016. He was awarded the Best Paper of ICDMA'15, and awarded an Outstanding Teacher of Software Design and Entrepreneurship Competition by the Ministry of Industry and Information Technology, China, in 2017. Dr. Ding was a recipient of the Medical Science and Technology Award (Second-Prize) of Jiangsu Province, China, in 2017, and the Education Teaching and Research Achievement Award (Third-Prize) of Jiangsu Province, China, in 2018. Dr. Ding was awarded two Chinese Government Scholarships for Overseas Studies in 2011 and 2016. Dr. Ding served /serves as an Associate Editor of *IEEE Transactions on Fuzzy Systems*, *Information Sciences*, *Swarm and Evolutionary Computation*, and *IEEE Access*.



Isaac Triguero received his M.Sc. and Ph.D. degrees in Computer Science from the University of Granada, Granada, Spain, in 2009 and 2014, respectively. He is currently an Assistant Professor in Data Science at the School of Computer Science of the University of Nottingham. He has published more than 30 international journal papers as well as more than 25 contributions to conferences. He currently serves as an associate editor of the Machine Learning and Knowledge Extraction journal, and the Big Data and Cognitive Computing journal. He is also a reviewer of more than 30 international journals. He has acted as Program Co-Chair of the IEEE Conference on Smart Data (2016), the IEEE Conference on Big Data Science and Engineering (2017), and the IEEE International Congress on Big Data (2018). He has acted as guest editor for special issues in journals such as Cognitive Computation, and Big Data Analytics. His research interests include data mining, data reduction, biometrics, optimization, evolutionary algorithms, semi-supervised learning, bioinformatics and big data learning.



Chin-Teng Lin (S'88-M'91-SM'99-F'05) received 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 the Distinguished Professor of Faculty of Engineering and Information Technology, University of Technology Sydney, University Chair Professor of Electrical and Computer Engineering, NCTU, International Faculty of University of California at San-Diego (UCSD), and Honorary Professorship of University of Nottingham. Dr. Lin was elevated to be an IEEE Fellow for his contributions to biologically inspired information systems in 2005, and was elevated International Fuzzy Systems Association (IFSA) Fellow in 2012. He was elected as the Editor-in-chief of *IEEE Transactions on Fuzzy Systems* in 2011-2016. He also served on the Board of Governors at IEEE Circuits and Systems (CAS) Society in 2005-2008, IEEE Systems, Man, Cybernetics (SMC) Society in 2003-2005, and IEEE Computational Intelligence Society (CIS) in 2008-2010. Dr. Lin is the Distinguished Lecturer of IEEE CIS Society from 2015-2017. He served as

the Deputy Editor-in-Chief of *IEEE Trans. on Circuits and Systems-II* in 2006-2008. He has published more than 200 journal papers (Total Citation: 19,166, H-index: 53, i10-index: 332) in the areas of fuzzy systems, neural networks and cognitive neuroengineering, including approximately 110 IEEE journal papers.