Extracting Multi-Knowledge from fMRI Data through Swarm-based Rough Set Reduction

Hongbo Liu[1], Ajith Abraham[2], Hong Ye[1]

[1] School of Computer Science, Dalian Maritime University, Dalian 116026, China
lhb@dlut.edu.cn, http://hongboliu.torrent.googlepages.com/
[2] Centre for Quantifiable Quality of Service in Communication Systems,
Norwegian University of Science and Technology, Trondheim, Norway
ajith.abraham@ieee.org, http://www.softcomputing.net/

Abstract. Functional Magnetic Resonance Imaging (fMRI) data is collected ceaselessly during brain research, which implicates some important information. It need to be extracted and translated to intelligible knowledge. In this paper, we attempt to extract multi-knowledge from fMRI data using rough set approach. A rough set reduction approach is presented based on particle swarm optimization algorithm, which discover the feature combinations in an efficient way to observe the change of positive region as the particles proceed through the search space. We illustrate some results using our approach, which is helpful for cognition research.

1 Introduction

Neuroinformatics is a research field that encompasses the development of neuroscience data and application of computational models and analytical tools. These areas of research are important for the integration and analysis of increasingly fine grain experimental data and for improving existing theories about nervous system and brain function [1]. Functional Magnetic Resonance Imaging (fMRI) is one of the most important tools for Neuroinformatics, which combines neuroscience and informatics science and computational science to develop approaches needed to understand human brain. Recently most of the research mainly arrange the activation features on region of interest (ROI) through statistical analysis, and neuroscientists or psychologists give out some explanation for the experimental results, which depends strongly on on their accumulative experience and subjective tendency. Rough set theory provides a novel approach to reduce the fMRI data and extracting knowledge. It helps us to derive rules from data represented in a given information system. The derivation of rules serves two main purposes: Firstly, the rules may be used in the classification of database objects, that is, predict the outcomes of unseen objects. Secondly, the rules may be used to develop a model of the domain under study, that is present knowledge in a format that can be understood by a human [2]. The rough set approach consists of several steps leading towards the final goal of generating rules from information/decision systems. The main steps of the rough set approach are: (1)

mapping of the information from the original database into the decision system format; (2) completion of data; (3) discretization of data; (4) computation of reducts from data; (5) derivation of rules from reducts; (6) filtering of rules. One of most important task is the data reduction process.

Conventional approaches always try to find a good reduct or to select a set of features [3]. In the knowledge discovery applications, only the good reduct can be applied to represent knowledge, which is called a single body of knowledge. In fact, many information systems in the real world have multiple reducts, and each reduct can be applied to generate a single body of knowledge. Therefore, multi-knowledge based on multiple reducts has the potential to improve knowledge representation and decision accuracy [4]. However, it would be exceedingly time-consuming to find multiple reducts in an instance information system with larger numbers of attributes and instances. In most of strategies, different reducts are obtained by changing the order of condition attributes and calculating the significance of different condition attribute combinations against decision attribute(s). It is a complex multi-restart processing about condition attribute increasing or decreasing in quantity. Swarm-based search approaches are of great benefits in the multiple reduction problems, because different individual trends to be encoded to different reduct. So it is attractive to find multiple reducts in the decision systems.

Particle swarm algorithm is inspired by social behavior patterns of organisms that live and interact within large groups [5]. The swarm intelligent model helps to find optimal regions of complex search spaces through interaction of individuals in a population of particles [6]. The particle swarm algorithm is particularly attractive for feature selection as there seems to be no heuristic that can guide search to the optimal minimal feature subset. Additionally, it can be the case that particles discover the best feature combinations as they proceed throughout the search space. This paper introduces particle swarm optimization algorithm to the rough set reduction processing and its application on fMRI data analysis.

2   Data Pre-processing

A typical normalized image contains more than 500,000 voxels, so it is impossible that feature vector can contain so immense voxels. We transform datasets from MNI template to Talairach coordinate system. Then we can use the region information in Talairach as features to reduce the dimensionality of the images. We used a SPM99 software package[3] and in-house programs for image processing, including corrections for head motion, normalization and global fMRI signal shift [7]. A simplified workflow is illustrated in Figure 1. Feature selection and extraction approach for fMRI data is composed of: (1) Find out the most active voxels in several regions of brain under the t-test of basic models in SPM99 and save their coordinates; (2) Scan fMRI image and search the voxels according to the coordinates saved; (3) Respectively average all voxels in the spherical region

---

[3] http://www.fil.ion.ucl.ac.uk/spm/

whose center is corresponding saved voxel and whose radius is a predefined constant. These results of a single image are formed one feature vector; (4) If the image is not the last one, go to Step 2, otherwise, end. The user interface for feature selection & extraction is illustrated in Figure 2.
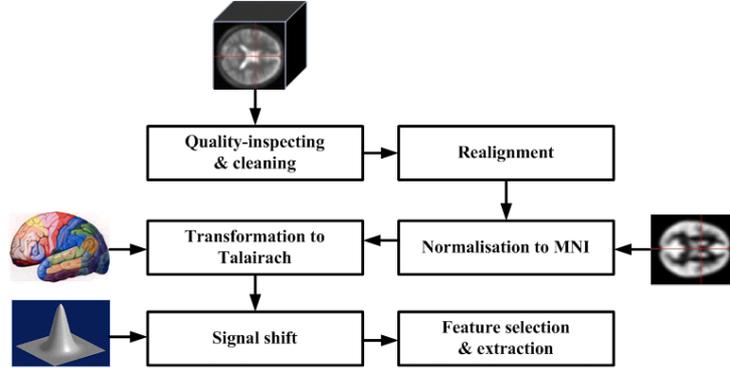

Fig. 1. Pre-precessing workflow for fMRI data
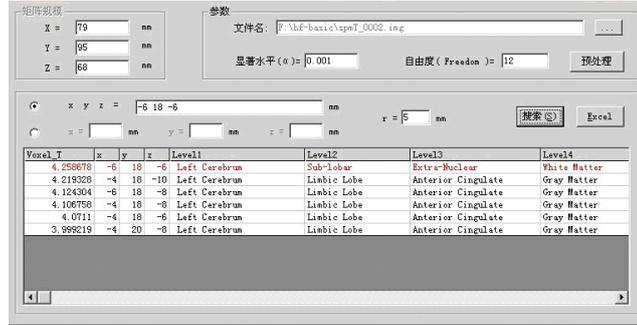
Fig. 2. Location for feature selection & extraction

3   Rough Set Reduction Algorithm

3.1   Reduction Criteria

The basic concepts of rough set theory and its philosophy are presented in [3, 8–10]. Here, we explain only the denotements relevant to our reduction method.

In rough set theory, an information system is denoted in 4-tuple by $S = (U, A, V, f)$, where U is the universe of discourse, a non-empty finite set of $|U|$ objects $\{x_1, x_2, \cdots, x_{|U|}\}$. A is a non-empty finite set of attributes such that $a : U \to V_a$ for every $a \in A$, where V_a is the value set of the attribute a and $V = \bigcup_{a \in A} V_a$. $f : U \times A \to V$ is the total decision function (also called the information function) such that $f(x, a) \in V_a$ for every $a \in A$, $x \in U$. The information system can also be defined as a decision table by $S = (U, C, D, V, f)$. For the decision table, C and D are condition and decision attributes, respectively.

Lower Approximation: Let $R \subseteq C \cup D$, $X \subseteq U$ and $U/R = \{R_1, R_2, \cdots, R_i, \cdots\}$. The R-lower approximation set of X is the set of all elements of U which can be with certainty classified as elements of X, assuming knowledge R. It can be presented formally as

$$APR_R^-(X) = \bigcup \{R_i | R_i \in U/R, R_i \subseteq X\} \tag{1}$$

Positive Region: Let $B \subseteq C$, $U/D = \{D_1, D_2, \cdots, D_i, \cdots\}$. The B-positive region of D is the set of all objects from the universe U which can be classified with certainty to classes of U/D employing features from B, i.e.,

$$POS_B(D) = \bigcup_{D_i \in U/D} APR_B^-(D_i) \tag{2}$$

Reduct: The attribute $a \in B \subseteq C$ is $D - dispensable$ in B, if $POS_B(D) = POS_{(B-\{a\})}(D)$; otherwise the attribute a is $D - indispensable$ in B. If all attributes $a \in B$ are $D - indispensable$ in B, then B will be called $D - independent$. A subset of attributes $B \subseteq C$ is a $D - reduct$ of C, iff $POS_B(D) = POS_C(D)$ and B is $D - independent$. Usually, there are many reducts in an instance information system. Let 2^A represent all possible attribute subsets $\{\{a_1\}, \cdots, \{a_{|A|}\}, \{a_1, a_2\}, \cdots, \{a_1, \cdots, a_{|A|}\}\}$. Let RED represent the set of reducts, i.e., $RED = \{B | POS_B(D) = POS_C(D), POS_{(B-\{a\})}(D) < POS_B(D)\}$.

Multi-knowledge: Let φ is a mapping from the condition space to the decision space. Then multi-knowledge can be defined as $\Psi = \{\varphi_B | B \in RED\}$.

Reduced Positive Universe: Let $U/C = \{[u_1']_C, [u_2']_C, \cdots, [u_m']_C\}$. Reduced Positive Universe U' can be written as:

$$U' = \{u_1', u_2', \cdots, u_m'\}. \tag{3}$$

and

$$POS_C(D) = [u_{i_1}']_C \cup [u_{i_2}']_C \cup \cdots \cup [u_{i_t}']_C. \tag{4}$$

Where $\forall u_{i_s}' \in U'$ and $|[u_{i_s}']_C/D| = 1$ $(s = 1, 2, \cdots, t)$.

Reduced Positive Region: Reduced positive universe can be written as:

$$U_{pos}' = \{u_{i_1}', u_{i_2}', \cdots, u_{i_t}'\}. \tag{5}$$

and $\forall B \subseteq C$, reduced positive region

$$POS_B'(D) = \bigcup_{X \in U'/B \wedge X \subseteq U_{pos}' \wedge |X/D| = 1} X \tag{6}$$

where $|X/D|$ represents the cardinality of the set X/D. $\forall B \subseteq C$, $POS_B(D) = POS_C(D)$ if $POS'_B = U'_{pos}$ [10]. It is to be noted that U' is the reduced universe, which usually would reduce significantly the scale of datasets. It provides a more efficient method to observe the change of positive region when we search the reducts. We only calculate U/C, U', U'_{pos}, POS'_B and then compare POS'_B with U'_{pos}.

3.2   Particle Swarm Approach for Reduction

The particle swarm model consists of a swarm of particles, which are initialized with a population of random candidate solutions. They move iteratively through the d-dimension problem space to search the new solutions, where the fitness f can be measured by calculating the number of condition attributes in the potential reduction solution. Each particle has a position represented by a position-vector $\boldsymbol{p}_i$ (i is the index of the particle), and a velocity represented by a velocity-vector $\boldsymbol{v}_i$. Each particle remembers its own best position so far in a vector $\boldsymbol{p}_i^{\#}$, and its j-th dimensional value is $p_{ij}^{\#}$. The best position-vector among the swarm so far is then stored in a vector $\boldsymbol{p}^*$, and its j-th dimensional value is p_j^*. The particle moves in a state space restricted to zero and one on each dimension. At each time step, each particle updates its velocity and moves to a new position according to Eqs.(7) and (8):

$$v_{ij}(t) = wv_{ij}(t-1) + \phi_1 r_1(p_{ij}^{\#}(t-1) - p_{ij}(t-1)) + \phi_2 r_2(p_j^*(t-1) - p_{ij}(t-1)). \quad (7)$$

$$p_{ij}(t) = \begin{cases} 1 & \text{if } \rho < sig(v_{ij}(t)); \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Where ϕ_1 and ϕ_2 are positive constants, r_1 and r_2 are the random numbers in the interval [0,1]. w is called as the inertia factor, ρ is random number in the closed interval [0,1].

Given a decision table $T = (U, C, D, V, f)$, the set of condition attributes, C, consist of m attributes. We set up a search space of m dimension for the reduction problem. Accordingly each particle's position is represented as a binary bit string of length m. Each dimension of the particle's position maps one condition attribute. The domain for each dimension is limited to 0 or 1. The value '1' means the corresponding attribute is selected while '0' not selected. During the search procedure, each individual is evaluated using the fitness. According to the definition of rough set reduct, the reduction solution must ensure the decision ability is the same as the primary decision table and the number of attributes in the feasible solution is kept as low as possible. In our algorithm, we first evaluate whether the potential reduction solution satisfies $POS'_E = U'_{pos}$ or not (E is the subset of attributes represented by the potential reduction solution). If it is a feasible solution, we calculate the number of '1' in it. The solution with the lowest number of '1' would be selected. The pseudo-code for the particle swarm search method is illustrated in Algorithm 1.

Algorithm 1 A Rough Set Reduct Algorithm Based on Particle Swarm

---

01.Calculate $U^{'}$, $U^{'}_{pos}$ using Eqs.(3) and (5).
02.Initialize the particle swarm, and other parameters.
03.While (the end criterion is not met) do
04. $t = t + 1$;
05.   Calculate the fitness value of each particle,
05.   if $POS^{'}_E \neq U^{'}_{pos}$, the fitness is punished
05.   as the total number of the condition attributes,
05.   else the fitness is the number of '1' in the position.
06. $\boldsymbol{p}^* = argmin^n_{i=1}(f(\boldsymbol{p}^*(t-1)), f(\boldsymbol{p}_1(t)), f(\boldsymbol{p}_2(t)), \cdots, f(\boldsymbol{p}_i(t)), \cdots, f(\boldsymbol{p}_n(t)));$
07.   For i= 1 to n
08. $\boldsymbol{p}^{\#}_i(t) = argmin^n_{i=1}(f(\boldsymbol{p}^{\#}_i(t-1)), f(\boldsymbol{p}_i(t)));$
09.   For $j = 1$ to d
10.   Update the j-th dimension value of $\boldsymbol{p}_i$ and $\boldsymbol{v}_i$
10.   according to Eqs.(7) and (8);
11.   Next j
12.   Next i
13.End While.

3.3   Experiments Using Benchmarks

To illustrate the effectiveness and performance of the considered algorithms, we illustrate the rough set reduct process and results through two benchmark problems. In this experiment, Genetic algorithm (GA) was used to compare the performance with PSO. In GA, the probability of crossover is set to 0.8 and the probability of mutation is set to 0.08. In PSO, ϕ_1 and ϕ_2 both are 1.49, and the inertia weight w is decreasing linearly from 0.9 to 0.1. The size of the population in GA and the swarm size in PSO both are set to $(even)(10+2*sqrt(D))$, where D is the dimension of the position, i.e., the number of condition attributes. In each trial, the maximum number of iterations is $(int)(0.1*recnum+10*(nfields-1))$, where $recnum$ is the number of records/rows and $nfields$ is the number of condition attributes. Each experiment (for each algorithm) was repeated 3 times with different random seeds. We consider two benchmark datasets, such as lung-cancer and lymphography from AFS[4].

Figure 3 illustrates the performance of the algorithms for lung-cancer and lymphography datasets, respectively. For lung-cancer dataset, the results (the number of reduced attributes) for 3 GA runs were 10: {1, 3, 9, 12, 33, 41, 44, 47, 54, 56} (The number before the colon is the number of condition attributes, the numbers in brackets are attribute index, which represents a reduction solution). The results of 3 PSO runs were 9: { 3, 8, 9, 12, 15, 35, 47, 54, 55}, 10: {2, 3, 12, 19, 25, 27, 30, 32, 40, 56}, 8: {11, 14, 24, 30, 42, 44, 45, 50}. For lymphography datasets, the results of 3 GA runs all were 7: {2, 6, 10, 13, 14, 17, 18}, the results of 3 PSO runs were 6: {2, 13, 14, 15, 16, 18}, 7: {1, 2, 13, 14, 15, 17, 18}, 7: {2, 10, 12, 13, 14, 15, 18}. PSO usually obtained a better result than GA.

---

[4] http://sra.itc.it/research/afs/
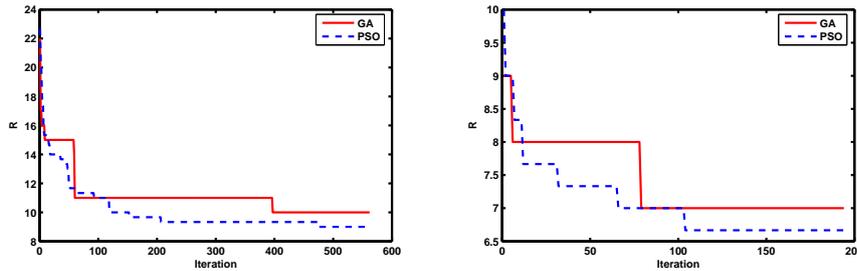
Fig. 3. Performance of rough set reduction for [a] lung-cancer dataset and [b] lymphography dataset

4   Extracting Knowledge from fMRI Data

We analyze the fMRI data from three cognition experiments: Tongue movement experiment, Associating Chinese verb experiment, and Looking at or silent reading Chinese word experiment. They are involved in 9 tasks: 0 - Control task; 1 - Tongue movement; 2 - Associating verb from single noun; 3 - Associating verb from single non-noun; 4 - Making verb before single word; 5 - Looking at number; 6 - Silent reading Number; 7 - Looking at Chinese word; 8 - Silent reading Chinese word. We select 13 ROIs and rank their actives as 0, 1, 2 and 3. The dataset includes 1 record for control task and 2580 records for other 8 cognitive tasks. In other words, the information system consists of 2581 rows and 14 columns. 2581 objects with 13 conditions and 1 decision attributes could be classified to 9 classes. The swarm size in PSO are set to 18. Each experiment (for each algorithm) was repeated 3 times with different random seeds. Other parameter settings are set to as same as ones for the benchmarks in last section. When our algorithm put out the reduction results, we check each particle for the solution. If the solution of the particle satisfies the criterion $POS_E^{'} = U_{pos}^{'}$, we accept the solution. It provided $18 * 3$ opportunities to extract the rules, although the length of these reduction results are not always "best", a minimum length. Some of rules are illustrated as follows:

Rule1: if M1=2, SMA=2, Broca=2 then Task=1;
Rule2: if BAs { 7,19,20,40,44,45 } =3, BSC=2 then Task=2;
Rule3: if BAs { 10,11,13,44,45 } =3, BSC=1 then Task=3;
Rule4: if BAs { 7,19,40 } =3, BSC=3 then Task=4;
Rule5: if SMA=2, Broca=3 then Task=6;
Rule6: if SMA=2, Broca=2, Wernike=3 then Task=8.

5   Conclusions

In this paper, we investigated the rough set approach to extract knowledge from fMRI Data. The data preprocessing workflow and mehtods for the rough set

approach were first discussed. A rough set reduction approach was presented based on particle swarm optimization algorithm, which discover the best feature combinations in an efficient way to observe the change of positive region as the particles proceed through the search space. We also evaluated the performance of particle swarm optimization algorithm with genetic algorithm based on two benchmark datasets. Empirical results indicated that PSO usually required shorter time to obtain better results than GA, although its stability need to be improved in further research. The swarm-based search approach provides some good benefits in the multiple reduction problems for multi-knowledge extraction. We illustrated some results using our approach for fMRI data analysis. Although the correctness of the rules need neuroscientists to analyze and verify further, the approach is helpful for cognition research.

Acknowledgments

References

1. Arbib, M. A., Grethe, J. S. (eds): Computing the Brain: A Guide to Neuroinformatics. Academic Press (2001)
2. Røed, G.: Knowledge Extraction from Process Data: A rough Set Approach to Data Mining on Time Series. Knowledge Extraction from Process data. Thesis, Norwegian University of Science and Technology (1999)
3. Wu, Q., Bell D.: Multi-knowledge Extraction and Application. Lecture Notes in Computer Science, 2639 (2003) 574–575
4. Wu, Q.: Multiknowledge and Computing Network Model for Decision Making and Localisation of Robots. Thesis, University of Ulster (2005)
5. Kennedy, J., Eberhart, R.: Swarm Intelligence, Morgan Kaufmann Publishers, San Francisco, CA (2001)
6. Clerc, M.: Particle Swarm Optimization. ISTE Publishing Company, London (2006)
7. Ji, Y., Liu, H., Wang X., Tang, Y.: Congitive States Classification from fMRI Data Using Support Vector Machines. In: Proceedings of The Third International Conference on Machine Learning and Cybernetics, IEEE Computer Society Press (2004) 2920–2923
8. Pawlak, Z.: Rough Sets and Intelligent Data Analysis. Information Sciences 147 (2002) 1–12
9. Wang, G.: Rough Reduction in Algebra View and Information View. International Journal of Intelligent Systems 18 (2003) 679–688
10. Xu, Z., Liu, Z., Yang, B., Song, W.: A Quick Attibute Reduction Algorithm with Complexity of $Max(O(|C||U|), O(|C|^2|U/C|))$. Chinese Journal of Computers 29 (2006) 391–399