Theoretical Research on Granular Computing and Artificial Selection Algorithm

Shoubai Xiao

Jiangxi University of Technology, Nanchang 330098, China

Keywords: Rough set; Granular computing; Binary granular matrix; Rough relationship matrix, Artificial Selection Algorithm.

Abstract. Today, with the rapid development of science and technology, knowledge and wisdom play greater roles and the evolution of artifacts is faster. The speed of artificial evolution cannot be followed by natural evolution. As a kind of new intelligent information processing technology, granular computing enjoys widespread concern of scholars. The constant integration of biological evolution, artificial life and calculation method has also been a hot research topic among scholars. The main findings of this paper include the following four parts: propose Rough set model based on granular computing; propose knowledge discovery algorithm based on granular computing; propose new project modeling method based on granular computing; propose Artificial Selection Algorithm (ASA) based on granular computing. This paper unifies the existing classic Rough set model, Rough set model based on probability and Rough set model based on inclusion degree all on the Rough set model based on granular computing.

Introduction

As can be seen from the development history of science, science was a unified system when it was in its infancy in the ancient civilization. From the late 19th century to early 20th century, science entered into a period of modern development with physics evolution as the beginning. With new ideas, new technologies and new means, people marched along with different levels of the material world, leading to the birth of emerging disciplines, inter-disciplines and integrated disciplines, and meanwhile, providing new fields, new angles and new ideas for scientific research. In the 20th century, genetics, gene theory and DNA double helix model were established, molecular biology was generated, and unprecedented achievements had been made in modern chemistry, biology, astronomy, cosmology and life science. The rise of information theory, cybernetics and system theory as well as the generation of self-organizing system theory provided a broader space for development of modern technology revolution. The 20th century became the era of atomic energy, the era of aerospace and the era of electronic technology and computer.

As time goes on, people gradually realize the importance of knowledge. The improvement of basic genetic algorithm is promoted to the improvement of natural evolution process based on knowledge. However, people's subjective initiative and data information produced in the evolution process have not been fully used. The role of humans in the evolutionary algorithm is small and passive.

As a classic and widely used evolutionary computation method, the genetic algorithm (GA) simulates the evolution law of "natural selection and survival of the fittest" in the natural evolution,

ignores the dynamic information accumulated in the evolution information, and completely separates itself from man-made interference in evolution. Though this interference is beneficial sometimes, it is the advantage of GA and also the defect of GA. Some scholars analyzed the mode, but the analysis was not sufficient.

From another perspective, the evolution process of GA is actually the process of knowledge production and accumulation. GA produces a large amount of knowledge in the evolution process. These data contains knowledge produced by the system in the evolution process. GA evaluates candidate solutions with fitness function, selects outstanding individuals in a random and probable mode, and approaches the optimal solution at a natural angle, but it does not clearly use the knowledge produced and accumulated by itself in the whole process. Therefore, a method of knowledge discovery can be adopted to dig out the original data. By reducing the dimensionality of data and complexity of data, the data is summarized, abstracted, refined and distinguished. Human wisdom is used and human requirements are considered to guide the evolution of GA.

According to this idea and based on standard GA, this paper introduces the concept of ASA. As for the research method, a new and effective tool granular computing idea for knowledge discovery and knowledge processing and the knowledge discovery method based on granular computing are used to dig out knowledge and refine law from data in the evolution process, exert the discovered knowledge in the evolutionary information system and simulate artificial selection process to lead evolution direction. Therefore, this paper focuses on description of granular computing, knowledge discovery algorithm based on granular computing and artificial selection algorithm based on granular computing.

Philosophical Foundation of Artificial Selection Algorithm

Since Ludwig von Bertalanffy founded the system theory, the development of this theory has been extremely rapid, showing its powerful vitality in application field. Meanwhile, the system perspective and system method it proposes has risen to the system philosophy.

The research object involved can be discussed and studied as a system to find out the inner certain relationship and reveal the general law, thus achieving the objective of recognizing and modifying the world. Therefore, the system theory is both outlook on world and methodology and has general applicability.

Marxism philosophy is the science about the most general laws of development of nature, society and human thinking. It focuses on study of scientific methodology, including the research on general method and specific method of natural science. The main methodologies include the outlook on nature and natural law of trinity of material, movement, space and time; scientific experiments and dialectical thinking; analysis and synthesis: hypothesis and proof. Obviously, natural science methodology is a combination point of natural sciences and philosophy.

Dialectical materialism holds that the world is material, the material is in motion, the motion is regular, and the law can be observed, used, recognized and mastered. By understanding and mastering the law of things, we can solve all kinds of problems better. We advocate combination of materialism and humanism because: our philosophy includes both ontology and creation; human creativity should start from objective reality and respect the objective laws. In any case, the matter is of great significance for human survival and development. We hope to achieve the unity of human and matter in the premise of people orientation.

With biological evolution theory as the theoretical dependence, genetic algorithm produces a huge number of historical data in its iterative evolution. In terms of the evolution of the system, the changing data is the most direct expression of physical movement as well as the carrier of the movement law. These huge data not only reflects the correlations between gene cluster and fitness in different evolution period, but also implies valuable information that directly reflects the law of evolution. There is law in the information hidden in individual iterative evolution. There is no doubt that these laws, once discovered, can be converted to heuristic knowledge, guiding and changing the original evolutionary trajectory meaningfully.

Biological science is also deeply influenced by the theoretical thinking and methods of philosophy. People should start from the objective reality and respect the objective laws during creative activities. The evolutionary process should be considered as a complex system. The knowledge discovery algorithm and statistical method can help us understand and master the law of genetic individuals and groups from junior to senior development and help us summarize genetic law for artificial selection and artificial breeding in accordance with human needs, thereby improving efficiency of traditional genetic algorithm.

Classic Rough Set Model

The concepts of rough set and common set are essentially different. Membership in rough set, and equivalence and inclusion of set are related to the universe knowledge expressed by the indiscernible relation of a set. Whether an element belongs to a set depends on people's knowledge rather than objectivity. So characteristic of rough set are not absolute but are related to the extent of our understanding of things. In this sense, the rough set approach can be seen as the expansion of classic set theory.

Classic Rough set model is also known as Pawlak Rough Set Model. It is based on the definition of approximate space to define the concepts such as upper and lower appropriate sets by introducing indiscernible relationships and to appropriately express knowledge from the perspective of rough level. From the perspective of the algorithm, upper and lower approximations can become two operators, collectively known as the rough operator. The upper and lower approximation sets are the results of calculation of these two operators.

Uncertainty and ambiguity of rough set theory are concepts based on boundaries. Each uncertain concept is expressed by the preciseness concept of a pair of upper approximation and lower approximation. Rough Set, equivalence relationship, upper approximation and lower approximation, boundary, positive field, negative field and indiscernible relationship are the basic concepts of rough set theory. Meanwhile, rough set theory gives the definitions of knowledge, knowledge base, knowledge representation system (information system) and decision-making table. The former is the cornerstone of rough set theory, while the latter forms the theoretical framework of knowledge discovery algorithm based on rough set theory.

In the classical rough set theory, all concepts and operations are defined according to algebraic equivalence relations and set operations, commonly known as algebraic expressions of rough set theory.

In Rough set theory, each uncertain concept is expressed by the preciseness concept of a pair of upper approximation and lower approximation. Border size indicates the level of uncertainty. Set

knowledge base K = (U, R). For any subset $X \subseteq U$ and an equivalence relation R, the set X can be divided according to description of R's basic set:

$$R_* X = U\{Y_t \in U \mid IND(R) : Y_i \subseteq X\}$$

$$R_* X = U\{Y_t \in U \mid IND(R) : Y_i \cap X \neq \phi\}$$

$$BN_R = R^* X - R_* X$$

$R_* X$和$R^* X$ are respectively called as R lower approximation and R upper approximation of x;

BN_R is called as X's border. R*X is the set of elements that are certainly included in X based on knowledge R. R*X is the set of elements that are certainly and possibly included in X based on knowledge R. When $BN_R \neq \phi$, X is an uncertain concept, and X is R rough set.

Rough Set Model Based on Probability

Based on the probability distribution, the concepts of information entropy and information quantity in information theory are introduced to Rough set to obtain the definition of information of Rough set as well as definition of information quantity, and meanwhile to provide new algorithms for knowledge simplicity.

Set U as the theory field constituted by finite objects and R is the equivalence relation on U. The constituted equivalence class is:

$$U / R = \{Y_1, Y_2 \ldots .Y_i \ldots .Y_m\} \, 1 \leq i \leq m$$

Set the equivalence class of x as Qin. Set P as the algebraic probability measure of σ constituted in the subset class of U. Triple group total A_p = (U, R, P) is called the probability approximation space. Each subset in U is called concept, representing a random event.

Upper approximation and lower approximation

Set $0 \leq \beta < \alpha \leq 1$, for any X⊂U, define A_p = (U, R, P). The lower approximation $PI_a(X)$ and upper approximation $\overline{PI}_\beta(X)$ of probability (I) according to parameter α, β are as follows:

$$\underline{PI}_\alpha(X) = \cup\{Y_i \in U \mid P(X \mid Y_i) \geq \alpha\}$$

$$\underline{PI}_\beta(X) = \cup\{Y_i \in U \mid P(X \mid Y_i) \geq \beta\}$$

$$POS(X, \alpha, \beta) = \underline{PI}_\alpha(X)$$

$$BN(X, \alpha, \beta) = U - \overline{PI_\beta(X)}^{()} - \underline{PI}_\alpha(X) = \cup\{Y_i \in U \mid \beta < P(X \mid Y_i) < \alpha\}$$

$$POS_X(X) = R_* X$$

$$neg(X, \alpha, \beta) = U - \overline{Pi}_\beta(X) = \cup\{Y_i \in U \mid P(X \mid Y_i) \le \beta\}$$

When a=1, =。 , the upper approximation and lower approximation of probability (I) are respectively upper approximation and lower approximation in Pawlak sense. Promotion form of probability rough set model-Pawlak rough set model.

Rough Set Model Based on Granular Computing

In the classical rough set theory, all the concepts and operations are defined according to algebraic equivalence relations and set operations, commonly known as algebraic expressions of rough set theory. In algebraic expression, the intuition of many concepts and operations of rough set theory is poor. It is not easy to understand its essence. In addition, in this expression, there is no efficient algorithm on knowledge simplicity.

Granular computing provides new ideas for us to solve problems. Meanwhile, rough set is seen as a special case of granular computing included in the theoretical framework of granular computing, providing new ideas for theoretical exploration and engineering applications of rough set theory.

The algorithm difficulties of the same question are different under different knowledge expressions.

$$U / IND(P) = \{Y_1, Y_2 \ldots . Y_i \ldots . Y_m\}; 1 \le i \le m$$ Given knowledge base K=(U, R). Given any

subset $P \subseteq R$, the equivalence class $U / IND(P) = \{Y_1, Y_2 \ldots . Y_i \ldots . Y_m\}; 1 \le i \le m$ is called as particle with knowledge P.

Use the binary number with length of l to define each particle with knowledge R. If $u_k \in Y_i, 1 \le k \le l,$ the i bit of binary number is 1; otherwise, it would be 0. Wherein, 1 is potential of U. Set the classes of p and Q attributes in theory field U are respectively Y and X.

$$Y = \{Y_1, Y_2 \ldots . Y_i \ldots . Y_m\} \qquad 1 \le i \le m$$
$$X = \{X_1, X_2 \ldots . X_j \ldots . X_n\} \qquad 1 \le j \le n$$

Carry out binary coding for particle $Y_i (\cup Y_i = U_i \cap Y_i = \phi)$ with knowledge P and particle $X_j (\cup X_j = U, \cap X_i = \phi)$ with knowledge Q.

$$Y_{m \times l} \triangleq \begin{pmatrix} a_{11} & \cdots & a_{1l} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{ml} \end{pmatrix}$$

$$X_{n \times l} \triangleq \begin{pmatrix} a_{11} & \cdots & a_{1l} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nl} \end{pmatrix}$$

$$C_{m \times n} = C_{YX} \triangleq Y \times X'$$

Matrix $C_{m \times n}$ reflects all inclusion relations between equivalence classes Y_i and X_j Therefore, we define $C_{m \times n}$ is the relationship matrix of P and Q

Conclusions

This paper establishes the Rough set model based on granular computing and proposes the knowledge discovery algorithm based on granular computing, engineering modeling algorithm based on granular computing, and artificial selection algorithm based on granular computing. The main conclusions are as follows:

1. This paper defines the binary granular matrix on the basis of binary particle and proposes the Rough set model based on granular computing. Meanwhile, this paper unifies all concepts and operations of rough set theory in this model, providing simple and intuitive methods to calculate the border of rough set.

2. Unify existing classical Rough set model, Rough set model based on probabilities and Rough set model based on inclusion together on the Rough set model based on granular computing to prove their equivalence.

3. Propose the concepts of certainty and coverage of decision-making algorithm based on granular computing.

4. The knowledge discovery algorithm based on granular computing is proposed. Through simple operation of binary granular matrix, this algorithm is able to deal with the decision-making compatible system as well as decision-making incompatible system, and achieve attribute reduction and property value reduction, providing a simple algorithm model for engineering application of granular computing. It proves that both the traditional algebraic reduction and information reduction in rough set can be attributed to numerical computation of granular matrix.

Acknowledgement

References

[1]The National Natural Science Fund Project (60374029) ''Theoretical and Technical Research on Rough-Fuzzy Intelligent Control Based on Thinking Evolution'' (Reference research)

[2]Shanxi Natural Science Fund (20051037) ''Image Processing Theory and Technology Based on Thinking Evolution-Rough Set'' (Reference research)

[3]Chen Zehua, Xie Keming. An Heuristic Artificial Selection Algorithm Based on Rough Set [J] Journal of Taiyuan University of Technology. 2007, 38 (1): 28-31

[4]Chen Zehua, Xie Keming, Xie Gang. Design and Simulation of Fuzzy Controller Based on Rough Set [J]. Journal of Electronics & Information Technology, 2003v (25): L-5

[5]Chen Zehua, Xie Gang, Xie Zeng, Xie Keming. Application of Binary Granular Matrix in Rough Set. Computer Science [J].

[6]Chen Zehua, Xie Gang, Xie Keming. Convergence Demonstration on Thinking Evolution Algorithm Based on Fixed Point Theory. Computer Science.2005, 32 (SA): 240-242

[7]Xie Keming, Tu Xinhong, Chen zehua. Basic Problems and Study of Granular Computing. Computer Engineering and Applications.

[8]Xie Keming, Chen Zehua, Xie Gang. Rough Set Theory and Its New Development in Intelligent Control[J]. Computer Science .2004 31 (IOA): 41-44

[9]Keming Xie，Zehua Chen， Gang Xie，T.Y Lin. BGrC for SuPerheated Steam TemPerature System Modeling in Power Plant[C]. Proc. Of 2006 IEEE International Conference on Granular Compution，Atlanta，USA， May10—12，2006:708—711

[10]Keming Xie，Zehua Chen.Fuzzy Forecas Modelingfor Gas Fumace Basedon Fuzzy Sets and Rough sets Theory[C] .Proe.10th Int. Conf.on Rough Sets，Fuzzy Sets， Data Mining，and Granular ComPuting(RSFDGrC 2005)，August 31st—SePtember 3rd，2005，Regina，Canada:614—623