1. Introduction

This document aims at summarizing my past research experience, expressing my research interest, and exploring the further research for which we can work together. I organize the content as the following structure: Firstly, I briefly review the representative papers I published in aspects of problem description, motivation, and the innovation from first-person perspective to enable you know my viewpoints to existing theories and algorithms, thinking process and scholar taste in research. Secondly, I would like to summarize my strength in application and leave some comments about the incorporating of granular computing into federated learning.

1. The prior research

This part is a detail explanation to my paper “improved general attribute reduction algorithms” from granular computing perspective. Firstly, I introduce basic concepts of Rough Sets in the way of information granule, and point out the key issues that were not resolved before this paper was published. Secondly, I explain the core of granularity space concepts proposed in the paper. In the final part, innovation and contribution are summarized.

## Basic concepts and problem existed in rough sets

The research object of rough sets is information system . Here stands for the universe of discourse, a non-empty finite set of instances. is the set of attributes. For a decision table, consists of condition attributes and decision attributes that satisfies and . is the set of all attribute values, and is an information function that maps an object in to exact on value in .

Rough sets attributes the proficiency of human’s problem-solving skills to the classification ability that is interpreted as reasonable concept in the universe. Information granule can be roughly described as multiple reasonable concepts. To get the information granule that requires minimal quantity of attributes and provides greatest classification performance, Pawlak proposed three approximations, i.e., positive region, negative region and boundary region, as main measurement for reasonability of information granules, and devised relative heuristic algorithms to compute a condition attribute set deducing target granule (called attribute reduct).

As mentioned above, the key issues in rough sets theory can be listed as:

* How to get reasonable information granules: In classic rough sets, there are five kinds of criteria at least, i.e., positive region preservation, generalized decision preservation, distribution preservation, maximum distribution preservation, distribution preservation, and relative discernibility relation preservation.
* Given a criterion or a measurement for information granule, how to get relative attribute sets deducing the target information granule: there are two mainstream methods: discernibility matrix-based reduction algorithms and heuristic reduction algorithms. In comparison to discernibility matrix-based reduction algorithms, heuristic reduction algorithms are more efficient but failed to get all attribute reducts.

There were problems in both theory and algorithms.

* Different but isolated criterion for evaluating information granule were proposed. This resulted in the increased theoretical complexity and perplexing definitions of attribute reducts, which is frustrating the application of rough sets.
* Almost all attribute reduction algorithms were low-efficiency that was partly caused by reliance of algorithms on isolated criterion. Different reduction algorithms were varied in the acceleration strategy.

## Information granule space

In this part, the problem “minimal cover set problem” is reviewed in comparison to the attribute reducts. Then, I interpret the attribute reduction as a variation problem of minimal cover set problem, which could be helpful to construction of a unified theory framework to cope with problems in rough sets theory.

Firstly, I would like to emphasize the similarity between minimal cover set problem and attribute reduct, which stirs up my interest on developing set-based definition of attribute reduct.

**Definition 1** (minimal cover set problem). given the universal set and subsets , a set of subsets was called the minimal cover set of if it satisfies:

1. There does not exist another set of subsets that meets the condition (1) and .

**Definition 2** (attribute reduct). Given a decision table , attribute sets , and a certain property of that can be represented by an evaluation function , an attribute set is called a reduct of if it satisfies following conditions:

It is obvious that two problems are very similar in the conditions. Now put it aside, we will turn back after going through the information granule of decision table.

In rough sets, information granules can be represented as partitions of the universe. For convenience, I prepare a simple example to help me to explain the concepts. For Table 1, we call any partition of as a granule, for instance, a partition deduced by attributes is .

Table 1. A decision table for indicating the information granules

|  |  |  |  |
| --- | --- | --- | --- |
| U | *a1* | *a2* | *d* |
| *x1* | 0 | 0 | 0 |
| *x2* | 1 | 0 | 1 |
| *x3* | 0 | 0 | 0 |

Information granules naturally give rise to hierarchical structures. For Table 1, the granule space can be depicted as Fig 1.

The interpretability and explain-ability of machine learning methods is a hot issue that until now there is no alternative way to completely solve it. I attribute the cause of this phenomenon into the lack of efficient tool to measure the uncertainty. The uncertainty can be grouped into two categories: data uncertainty and model uncertainty. For the data uncertainty, there is no unified theory or tool that can cope with noise and inconsistency. For the model uncertainty, we perform badly in answering what models really learn from data because our objective of learning always is quantity-sensitive, like variance of random variables that does not determine a unique probability distribution.

This paper “Improved general attribute reduction algorithm” provides an alternative answer to questions mentioned above when it comes to the classification task of discrete data. Before explaining the answer directly, we need to know main viewpoint of this paper.

The core concept of the paper is the granule space. Information granules naturally give rise to hierarchical structures. In rough sets, information granules can be described as partitions of the universe. If we have a universe of discourse and condition attributes (the decision table is placed at the left side of Fig 1), the granule space of this decision table can be depicted as the right part of Fig 1.

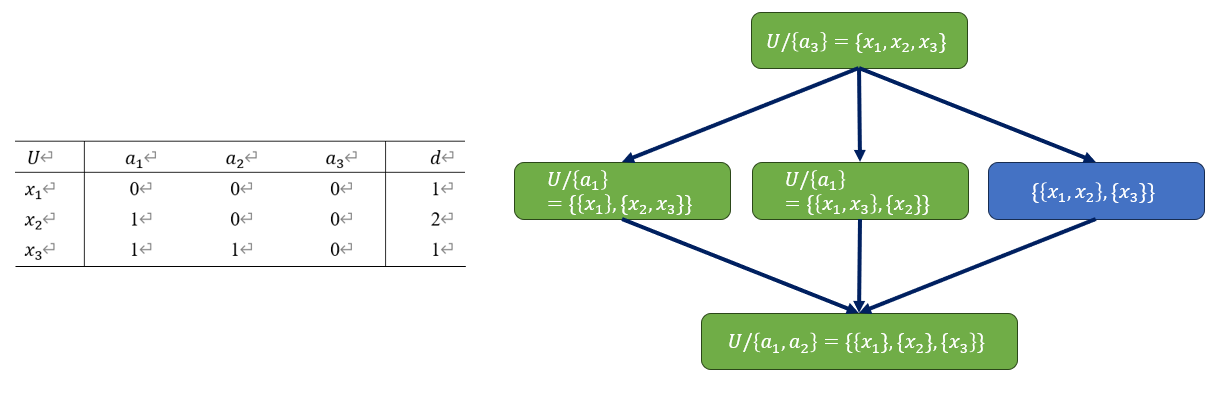


Fig 1. Granule space (right) of a decision table (left)

All rectangle items can be seen as the partition of and the green color denotes relative granule can be deduced by a set of attributes. Under the context of rough sets theory, the objective can be expressed as finding a partition or granule of that preserves the classification ability of whole decision table, and it can be deduced by a set of necessary attributes (necessary = elimination of any attribute in set can result in the recession of classification performance).

Let us consider a bit more complicated case, given a decision table with and , its granule space can be depicted as Fig 2. The innovation of my paper is the target granule that can determine the solution space. For example, if we want to train a machine learning model to fit or find an attribute reduct that guarantee the remains unchanged (distribution preservation), according to the computing method I proposed (CTGA algorithm in my paper), we get target granule . The ***granules of all attribute reducts*** or ***the granules machine learning models*** (no matter what machine learning methods you choose) ***learned*** ***must be one element of*** , which is depicted as Fig 3. Here *ANC* and *SPR* respectively denote ancestor granules and spring granules. If we have a path in granule space, we call *A* is an ancestor granule of *B* and *B* is a spring granule of *A*. In Fig 3, the grey items denote granules we don’t care if we want to get maximal classification performance; the other items are the fitting results or the qualified granules for preserving classification ability.

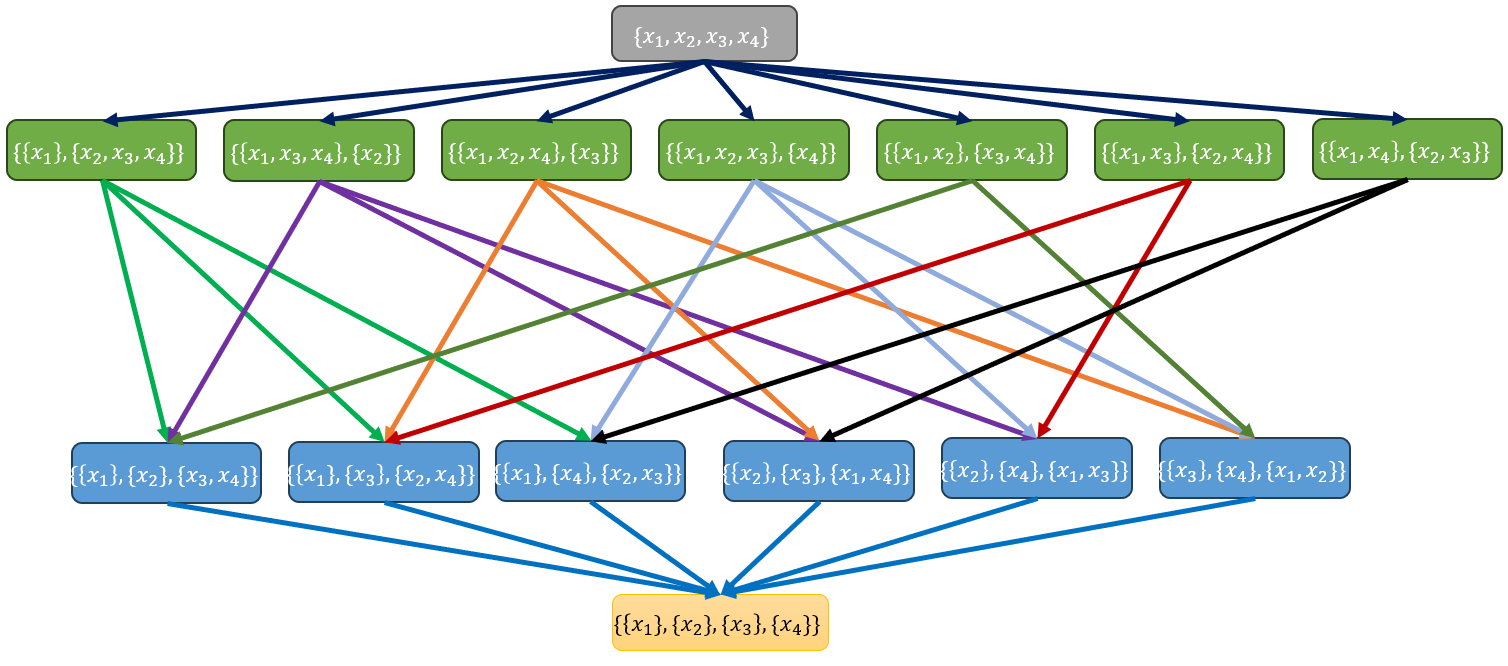


Fig 2. The granule space of the universe including four samples

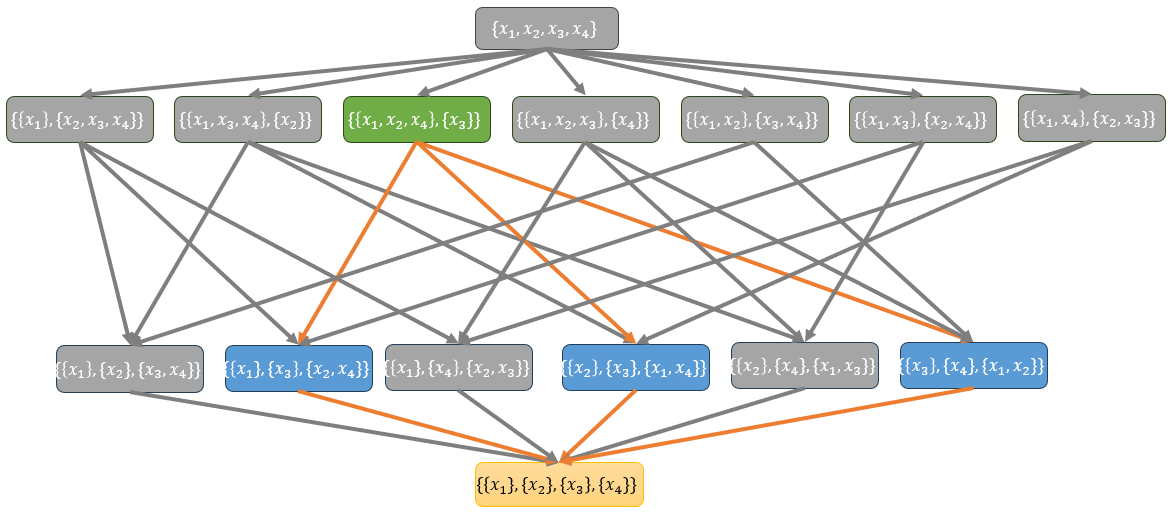


Fig 3. The granule space of fitting results or attribute reducts.

For the classification task of discrete data, granule space performs great in coping with data uncertainty and model uncertainty:

* For data uncertainty, the target granule is proposed as the refined representation of data, which can be incorporated with multiple strategies for the improvement of data quality, and it performs really good in coping with consistency and noise.
* Model uncertainty can be partly measured by the size of granules space determined by and . For the example of Fig 3, the relative granule space size is 5 and this means if you trained a model to fit , the information granule the model learned can be any one element of . These granules may have large variance and contributes to the model uncertainty. If we select only one granule in as the learning objective of models according to reasonable strategies, I believe this could dramatically decrease the model uncertainty and improve the performance of models.

As mentioned above, granule space is an efficient concept refining data and it has great potential to decrease the model uncertainty for discrete data, and I really want to extend this framework to the ordered data and continuous data in my future work.