Information Sciences 373 (2016) 308–320

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

# Rough data-deduction based on the upper approximation

Shuo Yan[a,*], Lin Yan[b], Jinzhao Wu[c,a]

[a] School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
[b] College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China
[c] Guangxi Key Laboratory of Hybrid Computation and IC Design Analysis, Guangxi University for Nationalities, Nanning 530006, China

ARTICLE INFO

*Article history:*
Received 24 December 2014
Revised 10 August 2016
Accepted 5 September 2016
Available online 5 September 2016

*Keywords:*
Rough deduction-space
Rough data-deduction
Deduction relation
Upper approximation
Rough relation
Support

ABSTRACT

This paper describes how to construct a structure called a rough deduction-space. It is an extension of an approximation space, and incorporates a deduction relation related to data connections. In the rough deduction-space, a notion of data deduction is introduced and is referred to as rough data-deduction. Based on integrated information of both the upper approximation and the deduction relation, rough data-deduction accomplishes deductions from data to data, which is different from any logical deduction in mathematical logic. Research on rough data-deduction covers two activities: the rough data-deduction with respect to an equivalence relation and rough data-deductions with respect to different equivalence relations. The activities also concern the relationship between rough data-deduction and rough relations that are rough representations of the deduction relation. This leads to properties involving approximate features implied by rough data-deduction, and reflecting the characteristic that rough data-deduction can describe rough data-connections. In particular, since the research correlates closely to granules, it may offer an avenue of research on granular computing. As an example, a specific problem is modeled by a rough deduction-space. The rough data-connections in the problem are described by use of rough data-deduction.



## 1. Introduction

The lower approximation and the upper approximation introduced in rough set theory [10] are important notions to approximate the representation of a concept, and are used to define rough sets. Moreover, research on rough sets has suggested different approaches [1–3,8,11,15,17–20,26,28–30] which demonstrate various processes of using them to deal with issues. At the same time, we noticed that the lower and upper approximations are always linked together in the approaches, and generally depend on each other in the process of representing a concept. However, when examining them separately, we can see that the lower approximation is limited to the inside of the granule that is the formal expression of the concept, whereas the upper approximation is generally an extension of the concept because it covers the granule. Therefore the lower approximation can be regarded as a rigorous representation, and the upper approximation can be viewed as a liberal representation about the concept.

Regarding the upper approximation, because it covers the granule, the upper approximation often contains the information which cannot be accurately characterized by the concept. We may as well refer to the information as approximate

* Corresponding author.
*E-mail addresses:* 11112097@bjtu.edu.cn (S. Yan), hnsdyl@163.com (L. Yan).

http://dx.doi.org/10.1016/j.ins.2016.09.011
0020-0255/

information which serves as our focus in the following discussion. On the other hand, because the lower approximation is limited to the inside of the granule, it needs to satisfy stronger conditions than those that the granule satisfies. The stronger conditions make the lower approximation too rigorous to have approximate information to be used. This motivates us to consider data processing that only correlates with the upper approximation, and does not concern the lower approximation.

In order to investigate the approximate information presented in the upper approximation, we intend to introduce a notion of data deduction. Data deduction will be based on the upper approximation, and will also incorporate information related to data connection. In fact, we want to integrate the approximate information into data deduction, so as to characterize unclear data connections by use of a deduction method. Unclear data connections are closely connected with actual problems that attract our attention. We are especially interested in a way to describe unclear data connections, and this serves as the focus of our paper. First, let us explore what is meant by unclear data connections.

**Example 1.** Let $A_1, A_2, \cdots, A_n (n \geq 2)$ be enterprises in an automobile manufacturing industry chain. They are linked together by supply and demand relations such that $A_i$ provides products to $A_{i+1} (i = 1, 2, \cdots, n-1)$. Then $A_1$, $A_2$, $\cdots$, and $A_n$ form a product supply channel, or show clear connections from $A_1$ to $A_n$ through $A_2$, $A_3$, $\cdots$, and $A_{n-1}$. Now consider $A'_1$ that is an enterprise of the same kind as $A_1$. In this case, there is a potential supply channel from $A'_1$ to $A_n$ because $A_1$ may be replaced by $A'_1$ in the production competition. Similarly, when $A'_n$ and $A_n$ are enterprises of the same kind, the potential connections from $A_1$ to $A'_n$ also exist. However, the connections from $A'_1$ to $A_n$ or from $A_1$ to $A'_n$ are not very clear.

Generally, if we observe things around us, we can be aware of the existence of unclear data connections with the same form as those in Example 1. Consider the following examples:

**Example 2.** If the persons $B_1$, $B_2$ and $B_3$ are drug traffickers, and $B_i$ sells drugs to $B_{i+1} (i = 1, 2)$, then there is a drug-dealing chain from $B_1$ to $B_3$ via $B_2$. If $B'_3$ is a partner or a friend of $B_3$, then there may be drug-dealing connections from $B_1$ to $B'_3$. However, the connections are not very clear.

**Example 3.** Consider the people $C_1$, $C_2$ and $C_3$, where $C_i$ lends money to $C_{i+1} (i = 1, 2)$. Thus the loan relations from $C_1$ to $C_3$ through $C_2$ are clear. Now if $C'_1$ is the son of $C_1$, then there may exist loan connections from $C'_1$ to $C_3$. However, the connections are not very clear.

**Example 4.** Let $D_1$, $D_2$ and $D_3$ be the people, such that $D_i$ is an immediate superior of $D_{i+1} (i = 1, 2)$. Obviously, the superior and subordinate relationships from $D_1$ to $D_3$ via $D_2$ are clear. Now if $D'_1$ is the wife of $D_1$, then there may be leadership connections from $D'_1$ to $D_3$. However, the connections are not very clear.

We can define these unclear data connections as *rough data-connections*. Thus, rough data-connections represent a variety of unclear data connections which have the same form as those in Example 1. Section 6 presents a formal discussion on the rough data-connections of Example 1, which will be based on our theoretical research.

Thus, some relationships between objects in actual problems are rough data-connections. In Example 1, the rough data-connections demonstrate the potential supply channel from $A'_1$ to $A_n$, or from $A_1$ to $A'_n$. The study of the rough data-connections may be valuable for enterprise management. Example 2 presents the rough data-connections as the indistinct drug-dealing connections from $B_1$ to $B'_3$. The connections may provide information for establishing a database, helpful for police in their efforts to combat drug abuse and drug trafficking.

So it is important to present a method to describe rough data-connections. The data deduction will be expected to be taken as the method. Indeed, we want to associate rough data-connections with deductions of data from data. This idea originates from mathematical logic [21] which focuses on the discussion of deduction among formulas. The formula deduction brings data deduction to our consideration because rough data-connections relevant to data instead of formulas can be considered as data deductions. We noticed that any deduction generally involves a direction that exists in rough data-connections as well. For instance, the potential supply channel from $A'_1$ to $A_n$ includes the direction from $A'_1$ to $A_n$; likewise, the indistinct drug-dealing connections from $B_1$ to $B'_3$ shows the direction from $B_1$ to $B'_3$, etc. This is consistent with a deduction process from antecedents to a consequent. On the other hand, some connections from data to data can broadly be viewed as data deductions. This also inspires us to consider the association between rough data-connection and a deduction method. Moreover, because rough data-connections are bound up with unclear information that reminds us of the approximate information in the upper approximation, this suggests that we may perhaps connect the data deduction with the upper approximation. In particular, in order to carry out data deductions, the information of data connection, such as data depend on data, data are associated with data, data can be derived from data, etc., also enters into our consideration.

All the ideas contribute to the research topic on data deduction. We will follow the ideas to develop our work. In this case, the work must be different from other research:

First, if we contemplate the research on rough sets together with the deductions, we can see that rough reasoning in rough logic is a way of combining logic and rough sets. The developments in [4,7,9,12,14,22–25,27] are the results of this study. However, the results are relevant to traditional methods used in mathematical logic. Rough reasoning still depends on formulas implementing formula deductions.

Second, from the researches into rough sets, such as those in [1–3,8,11,15,17–20,26,28–30], we know that the upper approximation is usually combined with the lower approximation to approximate a granule. According to the above discussion,