Learning a Table from a Table with Non-deterministic Information: A Perspective*

Mao Wu[1], Naoto Yamaguchi[1], Michinori Nakata[2], and Hiroshi Sakai[1]

[1] Faculty of Engineering, Kyushu Institute of Technology
Tobata, Kitakyushu 804, Japan
wumogaku@yahoo.co.jp, KITYN1124@gmail.com,
sakai@mns.kyutech.ac.jp

[2] Faculty of Management and Information Science
Josai International University
Gumyo, Togane, Chiba 283, Japan
nakatam@ieee.org

Abstract. *Rough Non-deterministic Information Analysis* (*RNIA*) is a rough sets-based framework for handling tables with exact and inexact data. In this framework, we have mainly investigated rough sets-based concepts in a table with non-deterministic information and some algorithms. This paper considers perspective on a new issue that how we estimate a table with actual information from a table with non-deterministic information by adding some constraint. This issue in *RNIA* slightly seems analogous to backpropagation in Neural Networks.

Keywords: Estimation of actual information, Constraint, Rough sets, Data dependency, Rules.

1 Introduction

Rough set theory offers a mathematical approach to vagueness [7]. It has many applications related to the areas of classification, feature reduction, rule generation, machine learning, data mining, knowledge discovery and others [8, 9]. Rough set theory is usually employed to deal with data tables with deterministic information, which we call *Deterministic Information Systems* (*DISs*). *Non-deterministic Information systems* (*NISs*) [5, 6] and *Incomplete Information systems* [3, 4] have also been investigated in order to handle information incompleteness.

We have been interested in *NISs*, and investigated *possible equivalence relations*, *data dependencies*, *rule generation*, *rule stability*, *question-answering systems*, as well as *missing* and *interval values* in *NISs* [10–16]. In each aspect,

* This work is supported by the Grant-in-Aid for Scientific Research (C) (No.22500204), Japan Society for the Promotion of Science.

A. Ell Hassanien et al. (Eds.): AMLTA 2012, CCIS 322, pp. 24–32, 2012.

modal concepts are employed, and each aspect is extended to two modes, namely the *certainty* and the *possibility*, or the *minimum* value and the *maximum* value.

In this paper, we describe perspective on new issue in *RNIA*. Namely, we consider methods that we estimate a table with actual information from a table with non-deterministic information by adding some constraint. This paper is organized as follows: Section 2 recalls the foundations of *RNIA*. Section 3 proposes to estimate a *DIS* from a *NIS* by consistency, dependency and rules. Section 4 concludes this paper.

2 Foundations of RNIA

A *Non-deterministic Information System* (*NIS*) Φ is a quadruplet [5–7].

$$\Phi = (OB, AT, \{VAL_A | A \in AT\}, g),$$
$$OB : \textit{ finite set whose elements are called objects},$$
$$AT : \textit{ a finite set whose elements are called attributes},$$
$$VAL_A : \textit{ a finite set whose elements are called attribute values},$$
$$g : OB \times AT \rightarrow P(\cup_{A \in AT} VAL_A)(\textit{a power set of } \cup_{A \in AT} VAL_A).$$

Every set $g(x, A)$ is interpreted as that there is an actual value in this set but this value is not known [5–7]. Especially if the real value is not known at all, $g(x, A)$ is equal to VAL_A. This is called the *null value* interpretation or missing value [3, 4]. We usually consider a table instead of this quadruplet Φ. Table 1 is an exemplary NIS Φ_1.

Table 1. An exemplary NIS Φ_1 for the suitcase data sets. Here, VAL_{Color}=$\{red, blue, green\}$, VAL_{Size}=$\{small, medium, large\}$, VAL_{Weight}=$\{light, heavy\}$, VAL_{Price}=$\{high, low\}$. In Φ_1, $g(x_1, Color)$=VAL_{Color}, and this means there is no information about this attribute value.

Object	*Color*	*Size*	*Weight*	*Price*
x_1	$\{red, blue, green\}$	$\{small\}$	$\{light, heavy\}$	$\{low\}$
x_2	$\{red\}$	$\{small, medium\}$	$\{light, heavy\}$	$\{high\}$
x_3	$\{red, blue\}$	$\{small, medium\}$	$\{light\}$	$\{high\}$
x_4	$\{red\}$	$\{medium\}$	$\{heavy\}$	$\{low, high\}$
x_5	$\{red\}$	$\{small, medium, large\}$	$\{heavy\}$	$\{high\}$
x_6	$\{blue, green\}$	$\{large\}$	$\{heavy\}$	$\{low, high\}$

Now, we introduce a *derived DIS* from a *NIS*, and show the basic chart in *RNIA*. Since each VAL_A ($A \in AT$) is finite, we can generate a *DIS* by replacing each non-deterministic information $g(x, A)$ with an element in $g(x, A)$. We named such a *DIS* a *derived DIS* from a *NIS*, and define the following.

$$DD(\Phi) = \{\psi \mid \psi \textit{ is a derived DIS from a NIS } \Phi\}.$$

In Φ_1, there are 2304 ($=3^2 \times 2^8$) derived $DISs$. The following DIS ψ_1 is an element of $DD(\Phi_1)$, namely $\psi_1 \in DD(\Phi_1)$ holds.

Table 2. A derived DIS ψ_1 from Φ_1

Object	*Color*	*Size*	*Weight*	*Price*
x_1	*red*	*small*	*light*	*low*
x_2	*red*	*small*	*light*	*high*
x_3	*red*	*small*	*light*	*high*
x_4	*red*	*medium*	*heavy*	*low*
x_5	*red*	*small*	*heavy*	*high*
x_6	*blue*	*large*	*heavy*	*low*

Due to the interpretation of non-deterministic information, we see an actual DIS ψ^{actual} exists in this 2304 derived $DISs$. Like this, we usually consider a set $DD(\Phi)$ of derived $DISs$ and the *basic chart* in Fig.1. We also coped with next modality.

(Certainty). If a formula α holds in every $\psi \in DD(\Phi)$, α also holds in ψ^{actual}. In this case, we say α *certainly holds* in ψ^{actual}.
(Possibility). If a formula α holds in some $\psi \in DD(\Phi)$, there exists such a possibility that α holds in ψ^{actual}. In this case, we say α *possibly holds* in ψ^{actual}.

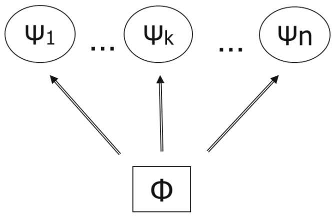

Fig. 1. An basic chart for Φ and a set $DD(\Phi)$ of derived $DISs$

Even if there exists the information incompleteness in Φ, we have the following decision making.

(1) If a formula α certainly holds, we think α holds under the uncertainty.
(2) If a formula α possibly holds, we think α may hold under the uncertainty.
(3) Otherwise, we think α does not hold under the uncertainty.

3 Learning a DIS from a NIS by Constraint

This section considers to estimate an actual DIS from a NIS.

3.1 A New Issue in RNIA

In the basic chart in Fig. 1, we considered $DD(\Phi)$ and defined the certainty and the possibility. We have already proposed some algorithms for handling them, and each algorithm is implicitly supposing some derived *DISs* for concluding two modalities. Namely, our previous work in *RNIA* took the following input and output.

(Previous Issue in *RNIA*)
Input: A *NIS*,
Output: Certain and possible conclusions with a set of supposed derived *DISs*.

In this paper, we consider the converse in output. Namely, we give constraint, and we estimate a set of supposed derived *DISs*.

(New Issue in *RNIA*)
Input: A *NIS*,
Output: A set of supposed derived *DISs* for concluding given constraint.

In each constraint, we have a set M_γ (γ: constraint) of derived *DISs*, and we will estimate an actual DIS as an element of $\cap_\gamma M_\gamma$. In the following subsections, we enumerate constraint, and intuitively explain each manipulation by using Φ_1.

3.2 Constraint 1: An Equivalence Class

A *possible equivalence class* X in $ATR \subset AT$ is a set of objects whose attribute values in each $A \in ATR$ are the same in a DIS $\psi \in DD(\Phi)$. Therefore, we are implicitly obtaining ψ for generating X. We take the converse, namely we define constraint γ (an equivalence class X), and then we have a set M_γ.

Example 1. In Table 3, if constraint γ is $X_{\{Color\}}=\{x_2, x_3, x_4, x_5\}$, the attribute value *red* (underlined in Table 3) is fixed in x_3. If constraint γ is that x_4 and x_6 do not belong to the same equivalence class in $Price$, we conclude either $x_4 : low$ and $x_6 : high$ or $x_4 : high$ and $x_6 : low$.

3.3 Constraint 2: Data Dependency

Data recovery by data dependency is known well. Functional dependency and data dependency are often employed for recovering missing values. We fix each attribute value which makes the degree of dependency [7] to take the maximum value.

Table 3. A part of Φ_1

Object	*Color*	*Price*
x_1	$\{red, blue, green\}$	$\{low\}$
x_2	$\{red\}$	$\{high\}$
x_3	$\{\underline{red}, blue\}$	$\{high\}$
x_4	$\{red\}$	$\{low, high\}$
x_5	$\{red\}$	$\{high\}$
x_6	$\{blue, green\}$	$\{low, high\}$

Example 2. In Table 4, if constraint γ is that there is data dependency from $Size \Rightarrow Price$. In this case, if we fix the following attribute values (underlined in Table 4),

$x_2 : [medium, high]$, $x_3 : [medium, high]$, $x_4 : [medium, high]$,
$x_5 : [\{medium, large\}, high]$,

there are three candidates of $DISs$ according to the values of x_5 and x_6. Namely, we fix the following in $\psi' \in DD(\Phi_1)$,

$x_5 : [medium, high]$, $x_6 : [large, \{low, high\}]$.

The other candidate of a DIS $\psi' \in DD(\Phi_1)$ is the following,

$x_5 : [large, high]$, $x_6 : [large, high]$.

In any case, the degree of dependency is 1.0 (=6/6). Like this in Φ_1, we can estimate three candidates of $DISs$ with actual information.

Table 4. A part of Φ_1

Object	*Size*	*Price*
x_1	$\{small\}$	$\{low\}$
x_2	$\{small, \underline{medium}\}$	$\{high\}$
x_3	$\{small, \underline{medium}\}$	$\{high\}$
x_4	$\{medium\}$	$\{low, \underline{high}\}$
x_5	$\{small, \underline{medium, large}\}$	$\{high\}$
x_6	$\{large\}$	$\{low, high\}$

In Example 2, this data set is very simple, and we could easily obtain the maximum degree of dependency from $Size \Rightarrow Price$. At first, we employed definite information in $x_1 : [small, low]$, then we fixed other attribute values. Namely, this procedure depends upon the validity of $x_1 : [small, low]$. However, this procedure may not be proper generally. In other cases, the ignorance of some definite information may make the degree maximum.

Generally, this calculation is translated to a combinatorial optimization problem, and the computational complexity is NP-hard [2]. Therefore, the calculation of the maximum degree for large data sets is not easy, and the estimation of attribute values may not be easy for large data sets, either.

3.4 Constraint 3: An Association Rule with Maximum Likelihood Estimation

In $RNIA$, we have proposed the following criteria for defining a rule τ.

(1) $minsupp(\tau)=Min_{\psi\in DD(\Phi)}\{support(\tau)\}$,
(2) $maxsupp(\tau)=Max_{\psi\in DD(\Phi)}\{support(\tau)\}$,
(3) $minacc(\tau)=Min_{\psi\in DD(\Phi)}\{accuracy(\tau)\}$,
(4) $maxacc(\tau)=Max_{\psi\in DD(\Phi)}\{accuracy(\tau)\}$.

Each criteria depends upon $DD(\Phi)$, and generally the number of $DD(\Phi)$ increases exponentially. Therefore, it will be difficult to calculate each criteria by enumerating each $\psi \in DD(\Phi)$. However, we have solved this problem by using two blocks inf and sup for each descriptor [13]. Furthermore, we have proved $minsupp(\tau)$ and $minacc(\tau)$ occur in the same derived DIS ψ_{min}. Similarly, $maxsupp(\tau)$ and $maxacc(\tau)$ occur in the same derived DIS ψ_{max}. Like this, we obtained the following chart [16].

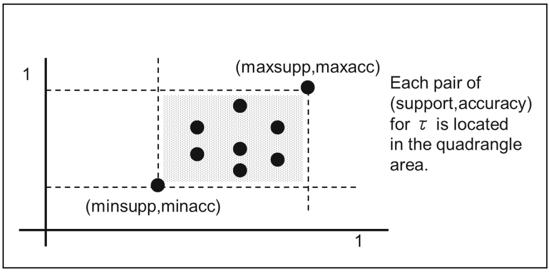

Fig. 2. A distribution of pairs ($support$,$accuracy$) for τ. There exists ψ_{min} which makes both $support(\tau)$ and $accuracy(\tau)$ the minimum. There exists ψ_{max} which makes both $support(\tau)$ and $accuracy(\tau)$ the maximum.

We also proposed *NIS-Apriori* algorithm by using inf and sup. *NIS-Apriori* algorithm implicitly handles $\psi \in DD(\Phi)$ for calculating four criterion values. We take the converse in *NIS-Apriori* algorithm, namely we give an association rule as a constraint and we estimate a set of $\psi_{max} \in DD(\Phi)$. We fix attribute values according to ψ_{max}, and this is the application of maximum likelihood estimation with constraint by an association rule.

Example 3. In Table 5, if constraint γ is that an association rule $[Color, red] \Rightarrow [Price, high]$ holds. Then, we fix attribute values for satisfying γ according to the *maximum likelihood estimation*, i.e.,

$x_1 : [\{blue, green\}, low]$, $x_3 : [red, high]$, $x_4 : [red, high]$,
$x_6 : [\{blue, green\}, \{low, high\}]$.

Furthermore, if the next constraint γ' is that an association rule $[Color, green] \Rightarrow [Price, low]$ does not hold. Then, we can reduce the attribute values to the following.

$x_1 : [blue, low]$, $x_3 : [red, high]$, $x_4 : [red, high]$,
$x_6 : [blue, \{low, high\}]$ or $[green, high]$.

Table 5. A part of Φ_1

Object	*Color*	*Price*
x_1	$\{red, \underline{blue, green}\}$	$\{low\}$
x_2	$\{red\}$	$\{high\}$
x_3	$\{\underline{red}, blue\}$	$\{high\}$
x_4	$\{red\}$	$\{low, \underline{high}\}$
x_5	$\{red\}$	$\{high\}$
x_6	$\{blue, green\}$	$\{low, high\}$

Since a specified association rule is valid in Φ_1, the procedure in Example 3 is always proper. Therefore, this constraint on an association rule is more convenient than the constraint on data dependency.

3.5 Constraint 4: Consistency

We have shown a set of constraint, and we have an intersection of $\cap_\gamma M_\gamma$. If $|\cap_\gamma M_\gamma| \geq 2$, we employ a strategy to keep consistency as much as possible, which we name *maximum consistency* strategy. It is possible to show an example to fix an attribute value. However, we are now considering the details of algorithms for this strategy.

3.6 An Example of Learning a DIS from a NIS

Now, we consider an exemplary NIS Φ_2 in Table 6.

Example 4. In Φ_2, we consider the first constraint

$\gamma_1 :$ *an association rule* $[Temperature, very\_high] \Rightarrow [Flu, yes]$,

and we employ *maximum likelihood estimation.* Then, we have the following attribute values are reduced.

$x_2 : [very\_high, yes, \{yes, no\}, yes]$, $x_3 : [very\_high, yes, yes, yes]$,
$x_8 : [very\_high, yes, \{yes, no\}, yes]$.

Table 6. An exemplary NIS Φ_2 for flu data sets. Here, $VAL_{Temperature}$=$\{normal, high, very\_high\}$, $VAL_{Headache}$=$\{yes, no\}$, VAL_{Nausea}=$\{yes, no\}$, VAL_{Flu}=$\{yes, no\}$.

$Object$	$Temperature$	$Headache$	$Nausea$	Flu
x_1	$\{high\}$	$\{yes, no\}$	$\{no\}$	$\{no\}$
x_2	$\{high, very\_high\}$	$\{yes\}$	$\{yes, no\}$	$\{yes\}$
x_3	$\{normal, high, very\_high\}$	$\{yes\}$	$\{yes\}$	$\{yes, no\}$
x_4	$\{high\}$	$\{yes\}$	$\{yes, no\}$	$\{yes, no\}$
x_5	$\{high\}$	$\{yes, no\}$	$\{yes\}$	$\{yes\}$
x_6	$\{normal\}$	$\{yes\}$	$\{yes, no\}$	$\{yes, no\}$
x_7	$\{normal\}$	$\{no\}$	$\{no\}$	$\{no\}$
x_8	$\{normal, high, very\_high\}$	$\{yes\}$	$\{yes, no\}$	$\{yes\}$

Table 7. An learned DIS ψ_2 from Φ_2

$Object$	$Temperature$	$Headache$	$Nausea$	Flu
x_1	*high*	*no*	*no*	*no*
x_2	*very_high*	*yes*	*yes*	*yes*
x_3	*very_high*	*yes*	*yes*	*yes*
x_4	*high*	*yes*	*yes*	*yes*
x_5	*high*	*yes*	*yes*	*yes*
x_6	*normal*	*yes*	*yes*	*yes*
x_7	*normal*	*no*	*no*	*no*
x_8	*very_high*	*yes*	*yes*	*yes*

We add the next constraint

γ_2 : *data dependency* $Headache \Rightarrow Nausea$,

then we have the following attribute values are reduced.

$x_1 : [high, no, no, no]$, $x_2 : [very\_high, yes, yes, yes]$,
$x_4 : [high, yes, yes, \{yes, no\}]$, $x_5 : [high, yes, yes, yes]$,
$x_6 : [normal, yes, yes, \{yes, no\}]$, $x_8 : [very\_high, yes, yes, yes]$.

Finally, we add the third constraint

γ_3 : *an association rule* $[Headache, yes] \wedge [Nausea, yes] \Rightarrow [Flu, yes]$.

Then, each attribute values in Φ_2 is uniquely fixed, and we have a DIS ψ_2 in Table 7 from Φ_2.

4 Concluding Remarks

This paper described how we estimate a DIS with actual information from a NIS. In $RNIA$, we tried to conclude the certainty and the possibility from a NIS, and implicitly we obtained a set of $DISs$ supporting the conclusion. We take the converse of this framework, namely we estimated a set M_γ of $DISs$ by constraint γ.

We have just started this work, and we are now investigating the manipulation for each constraint and the manipulation to estimate $\psi^{actual} \in \cap_\gamma M_\gamma$. Such manipulation seems analogous to backpropagation [17] in neural networks.

References

1. Aldrich, J.: R.A. Fisher and the making of maximum likelihood 1912-1922. Statistical Science 12(3), 162–176 (1997)
2. Cook, W.J., Cunningham, W.H., Pulleyblank, W.R., Schrijver, A.: Combinatorial optimization. John Wiley & Sons (1997)
3. Grzymała-Busse, J.W.: Data with Missing Attribute Values: Generalization of Indiscernibility Relation and Rule Induction. In: Peters, J.F., Skowron, A., Grzymała-Busse, J.W., Kostek, B.z., Świniarski, R.W., Szczuka, M.S. (eds.) Transactions on Rough Sets I. LNCS, vol. 3100, pp. 78–95. Springer, Heidelberg (2004)
4. Kryszkiewicz, M.: Rules in incomplete information systems. Information Sciences 113(3-4), 271–292 (1999)
5. Orłowska, E., Pawlak, Z.: Representation of nondeterministic information. Theoretical Computer Science 29(1-2), 27–39 (1984)
6. Pawlak, Z.: Systemy Informacyjne: Podstawy Teoretyczne. WNT (1983) (in Polish; English translation: Information Systems: Theoretical Foundations)
7. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers (1991)
8. Peters, G., Lingras, P., Ślęzak, D., Yao, Y. (eds.): Selected Methods and Applications of Rough Sets in Management and Engineering. Springer (2012)
9. Polkowski, L., Skowron, A. (eds.): Rough Sets in Knowledge Discovery, Parts 1 & 2. Physica-Verlag (1998)
10. RNIA software logs, `http://www.mns.kyutech.ac.jp/~sakai/RNIA`
11. Sakai, H., Okuma, A.: Basic Algorithms and Tools for Rough Non-deterministic Information Analysis. In: Peters, J.F., Skowron, A., Grzymała-Busse, J.W., Kostek, B.z., Świniarski, R.W., Szczuka, M.S. (eds.) Transactions on Rough Sets I. LNCS, vol. 3100, pp. 209–231. Springer, Heidelberg (2004)
12. Sakai, H.: Possible Equivalence Relations and Their Application to Hypothesis Generation in Non-deterministic Information Systems. In: Peters, J.F., Skowron, A., Dubois, D., Grzymała-Busse, J.W., Inuiguchi, M., Polkowski, L. (eds.) Transactions on Rough Sets II. LNCS, vol. 3135, pp. 82–106. Springer, Heidelberg (2004)
13. Sakai, H., Ishibashi, R., Koba, K., Nakata, M.: Rules and Apriori Algorithm in Non-deterministic Information Systems. In: Peters, J.F., Skowron, A., Rybiński, H. (eds.) Transactions on Rough Sets IX. LNCS, vol. 5390, pp. 328–350. Springer, Heidelberg (2008)
14. Sakai, H., Koba, K., Nakata, M.: Rough sets based rule generation from data with categorical and numerical values. Journal of Advanced Computational Intelligence and Intelligent Informatics 12(5), 426–434 (2008)
15. Sakai, H., Okuma, H., Nakata, M., Ślęzak, D.: Stable rule extraction and decision making in rough non-deterministic information analysis. International Journal of Hybrid Intelligent Systems 8(1), 41–57 (2011)
16. Sakai, H., Nakata, M., Ślęzak, D.: A Prototype System for Rule Generation in Lipski's Incomplete Information Databases. In: Kuznetsov, S.O., Ślęzak, D., Hepting, D.H., Mirkin, B.G. (eds.) RSFDGrC 2011. LNCS, vol. 6743, pp. 175–182. Springer, Heidelberg (2011)
17. Werbos, P.J.: The roots of backpropagation: From ordered derivatives to neural networks and political forecasting. John Wiley & Sons, New York (1994)