

1993

## Reconstructability Theory for General Systems and Its Application to Automated Rule Learning.

Sudhir Kumar Trivedi

*Louisiana State University and Agricultural & Mechanical College*

Follow this and additional works at: [https://digitalcommons.lsu.edu/gradschool\\_disstheses](https://digitalcommons.lsu.edu/gradschool_disstheses)

---

### Recommended Citation

Trivedi, Sudhir Kumar, "Reconstructability Theory for General Systems and Its Application to Automated Rule Learning." (1993). *LSU Historical Dissertations and Theses*. 5678.

[https://digitalcommons.lsu.edu/gradschool\\_disstheses/5678](https://digitalcommons.lsu.edu/gradschool_disstheses/5678)

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact [gradetd@lsu.edu](mailto:gradetd@lsu.edu).

## INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

# U·M·I

University Microfilms International  
A Bell & Howell Information Company  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
313/761-4700 800/521-0600



**Order Number 9419932**

**Reconstructability theory for general systems and its application  
to automated rule learning**

Trivedi, Sudhir Kumar, Ph.D.

The Louisiana State University and Agricultural and Mechanical Col., 1993

**U·M·I**  
300 N. Zeeb Rd.  
Ann Arbor, MI 48106



**RECONSTRUCTABILITY THEORY  
FOR GENERAL SYSTEMS  
AND ITS  
APPLICATION TO AUTOMATED RULE LEARNING**

**A Dissertation**

**Submitted to the Graduate Faculty of the  
Louisiana State University and  
Agricultural and Mechanical College  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy**

**in**

**The Department of Computer Science**

**by**

**Sudhir Kumar Trivedi  
B.Sc., Agra University, India, 1979  
M.Sc., Agra University, India, 1981  
Ph.D., Agra University, India, 1985  
December 1993**

*To Vicki*

## PREFACE

Reconstructability theory emphasizes the relationship between parts and wholes, the relationships between systems and subsystems, and more specifically, the relationship between states and substates. The two problems in the reconstructability theory are referred to as the reconstruction problem and the identification problem. The former relates to the process of reconstructing a given system under a given criterion from the knowledge of its subsystems and, during this process, identifying the subsystems that are important in the reconstruction. The latter allows the identification of an unknown system from the knowledge of its subsystems. The solution procedures associated with these two problems are referred to as Reconstructability Analysis, abbreviated as RA. Thus, the advent of RA has intensified the research efforts on system studies. The point of view is to consider the systems and subsystems as entities interrelated by some mathematical criterion. The objective of this research is to study the process of system reconstruction for general systems and to apply the results from RA to the problem of generalized rule induction from databases. We further the realm of RA by studying it in the context of probabilistic systems, selection systems and possibilistic systems. Based on RA methodology, we introduce a new paradigm to automated knowledge acquisition from databases. The following is a chapter outline of the dissertation.

Chapter 1 introduces the preliminary concepts in reconstructability theory and machine learning, such as systems, subsystems, states, substates, reconstructibility problem, identification problem, k-systems, automated knowledge acquisition, expert systems, and Baye's misclassification rate. Then, it briefly elaborates on the evaluation of reconstruction hypotheses, describes the significance of unbiased reconstruction, and explains the usefulness of k-systems and RA methodology.



Chapter 2 mainly modifies existing results in probabilistic RA. Previous work on reconstructability analysis for probabilistic systems and selection systems is extended to generate better algorithms for determining the unbiased reconstruction and reconstruction families. A proof for computing the unbiased reconstruction for U-structures is given when only independent information is employed.

Chapter 3 enhances the applicability of reconstructability theory by extending it for possibilistic systems using partial information. We introduce the concept of the partial reconstruction hypothesis, and compute the unbiased reconstruction and the reconstruction families implied by the partial information. A possibilistic version of the probabilistic algorithm is proposed to determine the unbiased reconstruction, and the reconstruction families have been identified by transforming the possibilistic system constraints to max-min fuzzy relation equations.

Chapter 4 applies the results of reconstructability analysis to the problem of automated knowledge acquisition from databases. When a large sample database exists, it is highly desirable to employ automated knowledge acquisition to learn important concepts. We introduce a new measure of the cognitive contents of a rule, called the K-measure. Based on reconstructability theory, our approach to rule learning from databases is unique in the sense that it should work for most data covered by the framework of RA. In particular, it is very appropriate for expert-systems-like domains where the data is intrinsically nominal. Unlike classical or classification-based techniques, reconstructability approach to automated rule learning does not induce any model on the data or on the nature of the solution, nor does it make any extraneous or erroneous assumptions regarding the data.

Chapter 5 provides concluding remarks. We summarize the results obtained and observe the potentials for further research in reconstructability analysis and learning. References are then ordered alphabetically in the bibliography.

Without inducing any order on these two names, Dr. J. Bush Jones and Dr. S. Sitharama Iyengar, I gratefully and respectfully state that it would not have been possible for me to complete my work without their generous and unequivocal support. I really am speechless while expressing my thanks to them. I must say I shall always remember them as role models in their respective modes.

My very special thanks go to my committee members, Dr. Donald H. Kraft, Dr. Jianhua Chen, Dr. Ishwar Murthy and Dr. Jianchao Zhu, who were always so understanding and supportive to me. Their very presence stimulated me to better myself. Further, I acknowledge that the part of my research was supported by the Office of Naval Research Grant Number ONR N0014-90K-0611.

Finally, to my roommates, Ram and Pankaj, who spent the nights with me typing and proofreading, I must tell that I would just not have made it in time without their support. I convey my sincere appreciation to them.

# TABLE OF CONTENTS

PREFACE .....	iii
ABSTRACT .....	viii
CHAPTER 1 INTRODUCTION .....	1
1.1 Preliminary Concepts in Reconstructability Theory .....	1
1.1.1 Systems and States .....	1
1.1.2 Subsystems and Substates .....	3
1.1.3 Reconstruction Problem .....	5
1.1.4 Identification Problem .....	6
1.2 Evaluation of Reconstruction Hypotheses .....	7
1.3 K-systems .....	9
1.4 Why RA and Why K-systems .....	10
1.5 Expert Systems and Knowledge Acquisition .....	11
CHAPTER 2 SOME NEW RESULTS IN RECONSTRUCTABILITY ANALYSIS .....	13
2.1 Determination of Reconstruction Families .....	13
2.1.1 Introduction .....	13
2.1.2 Determination of Triangular Matrix .....	15
2.1.2.1 Jones' Algorithm .....	15
2.1.2.2 A New Algorithm .....	16
2.1.3 Concluding Remarks .....	18
2.2 An Algorithm for Selection Functions .....	19
2.2.1 Introduction .....	19
2.2.2 Algorithm 1 .....	19
2.2.3 Algorithm 2 .....	20
2.2.4 Algorithm 3 .....	20
2.2.5 Concluding Remarks .....	23
2.3 Reconstruction of Probabilistic Behavior Functions .....	23
2.3.1 Introduction .....	24
2.3.2 Basic Join Procedure .....	24
2.3.3 Iterative Join Procedure .....	25
2.3.4 Unified Join Procedure .....	25
2.3.5 Concluding Remarks .....	28
2.4 Reconstructability Analysis for U-structures .....	29
2.4.1 U-algorithm .....	29
2.4.2 Concluding Remarks .....	31

<b>CHAPTER 3 RECONSTRUCTABILITY ANALYSIS OF POSSIBILISTIC SYSTEMS .....</b>	<b>32</b>
3.1 Introduction .....	32
3.2 Reconstruction of Possibilistic Systems .....	33
3.2.1 Reconstruction Algorithm .....	36
3.3 Reconstruction Families of Possibilistic Systems .....	39
3.3.1 Concluding Remarks .....	42
<b>CHAPTER 4 RECONSTRUCTABILITY APPROACH TO AUTOMATED RULE LEARNING .....</b>	<b>43</b>
4.1 Introduction .....	44
4.2 Other Approaches to Learning .....	45
4.3 Measures of Cognitive Contents of a Rule .....	48
4.4 KRULE: Rule Induction Algorithm .....	51
4.5 Advantages of KRULE Approach .....	53
4.6 Concluding Remarks .....	54
<b>CHAPTER 5 CONCLUDING REMARKS .....</b>	<b>55</b>
5.1 Summary and Conclusions .....	55
5.2 Problems and Potentials .....	57
<b>BIBLIOGRAPHY .....</b>	<b>61</b>
<b>VITA .....</b>	<b>66</b>

## ABSTRACT

The two problems in reconstructability analysis, abbreviated as RA, are referred to as the reconstructability problem and the identification problem. The former relates to the process of reconstructing a given system under a given criterion from the knowledge of its subsystems and, during this process, identifying those subsystems that are important in the reconstruction. The latter allows the identification of an unknown system from the knowledge of its subsystems. The advent of RA has intensified the research efforts on system studies. The objective of this research is to study the process of system reconstruction for general systems and apply it in the context of automated knowledge acquisition from databases.

First, we describe basic concepts in reconstructability theory and machine learning. We then modify existing results in reconstructability theory for probabilistic and selection systems in order to generate better algorithms for determining the unbiased reconstruction and reconstruction families in the wake of new developments such as k-systems and the use of independent information. Further, we extend RA methodology for possibilistic systems using only partial information. An algorithm is proposed to compute the unbiased reconstruction, and the reconstruction families are identified as a set of max-min fuzzy relation equations.

Furthermore, we define a new measure of the cognitive contents of a rule, referred to as the K-measure. Based on the K-measure, we introduce a new approach for automated knowledge acquisition from databases. Based on RA, the reconstructability approach to generalized rule induction from databases should work for most data covered by the framework of RA and k-systems. In particular, this approach is appropriate for expert-systems-like domains where the data is intrinsically nominal. Finally, we summarize our results and discuss the potentials for further research.

## CHAPTER 1

### INTRODUCTION

Reconstructability theory relates to two types of problems, namely, the reconstruction problem and the identification problem. The former deals with the process of reconstructing a given system under a given criterion from the knowledge of its subsystems, and during this process, identifying those subsystems that are instrumental in the reconstruction. The latter allows the identification of an unknown system from the knowledge of its subsystems. The solution procedures associated with these two problems are referred to as Reconstructability Analysis, abbreviated as RA. Origins of RA can be traced to Ashby's work on constraint analysis in the early sixties [2, 47], though a formal framework of RA did not exist until the late seventies [7, 43, 44] and early eighties [8-11]. Since the advent of reconstructability theory, significant efforts have been directed towards research in this area, resulting in the emergence of a variety of new algorithms and applications. Solution procedures aimed at these two problems have been developed and implemented [26-27, 33-41, 43-48, 54-55, 66]. In this chapter we shall introduce the basic terminology and concepts in reconstructability theory and learning [9, 64].

#### 1.1 Preliminary Concepts in Reconstructability Theory

##### 1.1.1 Systems and States

Intuitively, a system is simply a data set which consists of the tuples of the form  $\langle v_1, v_2, \dots, v_n, f \rangle$ , where  $v_1, v_2, \dots, v_n$  are variables or attributes and  $f$  is a function defined over these variables. This function may be a probability distribution function, a possibilistic behavior function, a selection function, a fuzzy set membership function or any arbitrary or non-linear function. A state in a system is simply a combination of

the attribute values in a given order. Formally, a system is defined as a six-tuple

$$B = (V, W, s, A, Q, f) \quad (1.1)$$

where  $V = \{v_i | i \in 1, 2, \dots, n\}$  is a set of variables;  $W = \{V_j | j \in \{1, 2, \dots, m\}, m \leq n\}$  is a family of state sets;  $s: V \rightarrow W$  is an onto mapping which assigns to each variable in  $V$  one state set from  $W$ ;  $A = s(v_1) \times s(v_2) \times \dots \times s(v_n)$  is the set of all potential aggregate states;  $Q$  is a set of real numbers; and  $f: A \rightarrow Q$  is a system function which represents the information regarding the aggregate states of the system. Note that we have generalized the meaning of the terms  $B$  and  $f$  as we refer to them as system and system function, respectively, instead of behavior system and behavior function as was originally used in [9]. This generalization is evident from the fact that RA methodology has evolved to cover functions other than behavior functions [38-41]. Following is an example of a probabilistic behavior function.

$\overline{v_1}$	$\overline{v_2}$	$\overline{v_3}$	$\overline{f}$
0	0	0	0.079
0	0	1	0.088
0	0	2	0.083
0	1	0	0.031
0	1	1	0.052
0	1	2	0.097
1	0	0	0.091
1	0	1	0.072
1	0	2	0.037
1	1	0	0.109
1	1	1	0.128
1	1	2	0.133

In the above example  $V = \{v_1, v_2, v_3\}$  is the set of variables;  $W = \{\{0, 1\}, \{0, 1, 2\}\}$  is a family of state sets;  $s: V \rightarrow W$  is an onto mapping which assigns  $\{0, 1\}$  to  $v_1$  and  $v_2$ , and  $\{0, 1, 2\}$  to  $v_3$ ;  $A = \{000, 001, 002, 010, 011, 012, 100, 101, 102, 110, 111, 112\}$ ;  $Q = [0, 1]$  is a set of real numbers and  $f: A \rightarrow Q$  is a probabilistic behavior function.

### 1.1.2 Subsystems and Substates

A subsystem is a data set whose variables form a proper subset of the variables of the system and a function  $g$  is defined over variables in the subset. It consists of the tuples of the form  $\langle u_1, u_2, \dots, u_m, g \rangle$ , where  $\{u_1, u_2, \dots, u_m\} \subset \{v_1, v_2, \dots, v_n\}$ . A substate in a subsystem is a combination of attribute values present in that subsystem in a given order. The concept of a system and subsystem is a relative one. A system can be regarded as a subsystem of a larger system and a subsystem can be regarded as a system (i.e., a supersystem) of a smaller system.

Formally, given a system as defined above, a collection of a total of  $q$  subsystems, together called a structure system or a reconstruction hypothesis, is defined as

$$S = \{ {}^k B \} = \{ ( {}^k V, {}^k W, {}^k s, {}^k A, {}^k Q, {}^k f ) \mid k \in \{1, 2, \dots, q\} \} \quad (1.2)$$

if and only if, for each  $k$ , following conditions are satisfied:

- (1)  ${}^k V \subset V$ ,
- (2)  ${}^k W \subseteq W$  such that  ${}^k s$  is onto,
- (3)  ${}^k s: {}^k V \rightarrow {}^k W$  such that  ${}^k s(v_i) = s(v_i)$  for each  $v_i \in {}^k V$ ,
- (4)  ${}^k A = \times_{v_i \in {}^k V} {}^k s(v_i)$ ,
- (5)  ${}^k Q = Q$ ,
- (6)  ${}^k f = [f \downarrow {}^k V]$ .



Elements of set  $S$  are referred to as subsystems of system  $B$ .  $[f \downarrow {}^k V]$  is called the projection of  $f$  on  ${}^k V$ , which considers only the variables in  ${}^k V$ . Essentially,  $[f \downarrow {}^k V]$  is a mapping from substates in  ${}^k A$  to  $Q$ , that is,

$$[f \downarrow {}^k V]: \times_{v_i \in {}^k V} s(v_i) \rightarrow Q \quad (1.3a)$$

such that

$$[f \downarrow {}^k V](\beta) = g(\{f(\alpha) \mid \alpha > \beta\}), \quad (1.3b)$$

where  $\alpha > \beta$  means  $\beta$  is a substate of  $\alpha$  (or  $\alpha$  is a superstate of  $\beta$ ), and  $g$  is determined by the nature of function  $f$ . For instance, if  $f$  is a probabilistic distribution function, then

$${}^k f(\beta) = \sum_{\alpha > \beta} f(\alpha), \quad (1.4a)$$

or if  $f$  is a selection or possibilistic behavior function, then

$${}^k f(\beta) = \max_{\alpha > \beta} f(\alpha). \quad (1.4b)$$

Following is an example of typical subsystems of the previously defined system.

$v_1$	$v_2$	${}^{12}f$	$v_2$	$v_3$	${}^{23}f$	$v_1$	$v_3$	${}^{13}f$
—	—	—	—	—	—	—	—	—
0	0	0.25	0	0	0.17	0	0	0.11
0	1	0.18	0	1	0.16	0	1	0.14
1	0	0.20	0	2	0.12	0	2	0.18
1	1	0.37	1	0	0.14	1	0	0.20
			1	1	0.18	1	1	0.20
			1	2	0.23	1	2	0.17

$v_1$	$^1f$	$v_2$	$^2f$	$v_3$	$^3f$
—	—	—	—	—	—
0	0.43	0	0.45	0	0.31
1	0.57	1	0.55	1	0.34
				2	0.35

In the above example, the reconstruction hypothesis can be described as  $S = \{^k B\} = \{\{v_1, v_2\}, \{v_1, v_3\}, \{v_2, v_3\}, \{v_1\}, \{v_2\}, \{v_3\}\}$ . Illustratively,  $^{12}V = \{v_1, v_2\}$  is the set of variables;  $^{12}W = \{\{0, 1\}\}$  is a family of state sets;  $^{12}s: ^{12}V \rightarrow ^{12}W$  is an onto mapping which assigns  $\{0, 1\}$  to  $v_1$  and  $v_2$ ;  $^{12}A = \{^{12}(00), ^{12}(01), ^{12}(10), ^{12}(11)\}$ ;  $^{12}Q = [0, 1]$  is a set of real numbers; and  $^{12}f: ^{12}A \rightarrow ^{12}Q$  is a probabilistic behavior function. Similarly,  $^3V = \{v_3\}$  is the set of variables;  $^3W = \{\{0, 1, 2\}\}$  is a family of state sets;  $^3s: ^3V \rightarrow ^3W$  is an onto mapping which assigns  $\{0, 1, 2\}$  to  $v_3$ ;  $^3A = \{^3(0), ^3(1), ^3(2)\}$ ;  $^3Q = [0, 1]$  is a set of real numbers; and  $^3f: ^3A \rightarrow ^3Q$  is a probabilistic behavior function. It should be noted that the above subsystems are derived using projection functions. This may not always occur in the real world where different subsystems may be observed by different teams of observers or by using different experiments. This gives rise to the issues of local and global inconsistencies, which we shall discuss later.

### 1.1.3 Reconstructability Problem

Let  $B$  be a behavior system defined by (1.1). Let  $S$  be a structure system defined by (1.2).  $S$  is said to be a meaningful reconstruction hypothesis of  $B$  if and only if it contains the subsystems of  $B$  such that

$$\bigcup_{k \in N_q} {}^k V = V, \text{ and} \quad (1.5)$$

$$(\text{ for all } j, k \in N_q) ({}^j V \subseteq {}^k V \Rightarrow j = k). \quad (1.6)$$

Condition (1.5) is called the *covering condition* and guarantees that all variables of  $B$  are included in  $S$ . This means that the reconstruction of  $B$  from  $S$  is logically possible. Condition (1.6) is called the *irredundancy condition* and ensures that  $S$  contains no redundant information.

#### 1.1.4 Identification Problem

Let  $S$  be a structure system defined by (1.2). Then  $S$  is said to be a reconstruction hypothesis (hypothetical representation) of an unknown overall system  $B$  provided the following six conditions hold true:

- (1)  $V = \bigcup_{k \in N_q} {}^k V$ ,
- (2)  $W = \bigcup_{k \in N_q} {}^k W$ ,
- (3)  $s: V \rightarrow W$  such that  $s(v_i) = {}^k s(v_i)$  for each  $k \in N_k$ ,
- (4)  $A = \times_{v_i \in V} s(v_i)$ ,
- (5)  $Q = {}^k Q$  for each  $k \in N_q$ ,
- (6)  $f: A \rightarrow Q$  such that  $[f \downarrow {}^k V] = {}^k f$  for each  $k \in N_q$ .

It is obvious that  $B$ , which is unknown, should be compatible with  $S$ . Potentially, there will be more than one overall systems compatible with  $S$ . The set of all these systems is called *reconstruction family* of  $S$ , denoted by  $B_S$ .

As discussed previously, if the behavior functions  $\{{}^k f\}$  of a reconstruction hypothesis are projections of an overall behavior function  $f$ , then the reconstruction hypothesis is *consistent*. A reconstruction hypothesis is said to be locally consistent if the following condition, called *local consistency condition*, is satisfied:

$$(\text{ for all } j, k \in N_q) ([{}^j V \downarrow {}^j V \cap {}^k V] = [{}^k V \downarrow {}^j V \cap {}^k V]). \quad (1.7)$$

A reconstruction hypothesis  $S$  is said to be *globally consistent* if the reconstruction family of  $S$  is non-empty. It is usually the case that the reconstruction hypotheses satisfy local consistency condition (1.7) as well as the covering condition (1.5) and the irredundancy condition (1.6).

## 1.2 Evaluation of Reconstruction Hypotheses

Let  $B$  be a given system [9]. Then there exists a family of meaningful reconstruction hypotheses  $\zeta_B$ , for  $B$ . Given a particular reconstruction hypothesis  $S$  in this family, there exists  $B_S$ , a family of systems which are compatible with the hypothesis. By being compatible we mean that the projections of each system function are those subsystem functions that are included in that particular reconstruction hypothesis.  $B_S$  is called the reconstruction family of  $S$ . The value of a suitable measure, referred to as reconstruction uncertainty, is used to express the size of the reconstruction family. This, in turn, defines an identifiability quotient, a measure of identifying a unique system from a given reconstruction hypothesis. The task to make a choice of a single system function  $f_S$  from the reconstruction family  $B_S$  requires some assumptions in order to justify the particular choice. The most significant theoretical justification [9] is that the determined system function  $f_S$  should be non-committal in all regards except in satisfying the following condition,

$$[f_S \downarrow^k V] = {}^k f = [f \downarrow^k V] \quad (1.8)$$

for all  $k \in \{1, 2, \dots, q\}$ . For a probabilistic system function  $f$ , the above condition implies that the set of values  $\{f_S(\alpha) \mid \alpha \in A\}$  must have the maximum entropy among all such sets associated with the systems in  $B_S$ . This implies that the reconstruction must be a maximum entropy solution subject to condition (1.8). In the context of a reconstruction problem, as stated in [9], the principle of maximum entropy can be justified by the following arguments:

- (i) The maximum entropy reconstruction is the only unbiased reconstruction as it takes into account all available information but no additional information [28-29].
- (ii) The maximum entropy reconstruction is the most likely reconstruction. For instance, if we are given a reconstruction hypothesis, each system in  $B_S$  can be generated by any number of actual data sets. The largest number of possible data sets that are compatible with  $S$  are those that are also compatible with the maximum entropy reconstruction [28-29].
- (iii) Maximizing any function other than entropy will lead to inconsistencies except when that function has the same maxima as entropy [63].
- (iv) Every real world system can be represented by the maximum entropy reconstruction. In other words, if we know that a system is a real world system, the maximum entropy reconstruction is the only one possible. It is impossible to design a real world system whose actual reconstruction is different from the maximum entropy reconstruction [9].

For probabilistic systems, Shannon's entropy is used as the measure of uncertainty which satisfies the properties of symmetry, expansibility, subadditivity, additivity, normalization and continuity [45, 62]. For a possibilistic system, Shannon's entropy is replaced by *U-uncertainty* [26, 45] which satisfies not only the properties satisfied by its probabilistic counterpart but also the property of monotonicity. Therefore, arguments (i) to (iv) also hold good to justify the principle of maximum uncertainty reconstruction in the context of possibilistic systems [11, 27, 48].

The problem of determining the unbiased reconstruction  $f_S$  for a reconstruction hypothesis  $S$  can be formulated as an optimization problem [9]. Given the set of functions  $h$  such that  $h: A \rightarrow Q$ , determine  $f_S$  for which the entropy  $-\sum_{\alpha \in A} h(\alpha) \log h(\alpha)$  reaches its maximum subject to condition (1.8). Cavallo and Klir [9] provide a

solution to this problem. Later, Cavallo and Klir [11], Higashi, et. al. [27] and Klir, et. al. [48] extend RA to cover the possibilistic structure systems and possibilistic system functions using the principle of maximum uncertainty.

Motivated by Lewis' study [50] on approximation of probability distributions to reduce storage requirements, Jones [33-35] provides the alternative methods of solution using minimal and limited information. He introduced the concept of null extension and k-system theory, and extended the previous results for the case when system functions were allowed to be any non-linear function, and this was accomplished only by means of the limited independent information [33-38]. He further generalized these results to hold good for incomplete and arbitrary data [39]. The concept of null extension and the advent of k-theory has greatly extended the realm of RA methodology. By transforming any non-linear system to a dimensionless system, it is possible for RA to cover most non-linear functions. Using the concept of null extension, it is possible to divide the whole state space into disjoint equivalent classes and to proceed further by just picking only one state from each class.

### 1.3 K-systems

Jones has extended the system/subsystem paradigm of RA to the state/substate paradigm by modifying RA to work at the levels of states and substates [33-41], and to work for general non-linear functions [38]. He has formalized the concepts of a g-system and a k-system. Let  $R^+$  be the set of positive real numbers. Let  $f$  and  $\tau$  be defined as follows:

$$f: A \rightarrow R^+, \quad (1.9a)$$

$$\tau = \sum_{\alpha \in A} f(\alpha). \quad (1.9b)$$

Then a g-system is defined as the following tuple:

$$(\tau, \{v_i\}, \{\alpha\}, \{\beta\}, f, \{^m f\}). \quad (1.10)$$

Here  $\tau$  is a parameter of interest to define k-systems,  $\{v_i\}$  is a set of variables,  $\{\alpha\}$  is a set of states,  $\{\beta\}$  is a set of substates,  $f$  is a function on  $\{\alpha\}$ , and  $\{^m f\}$  are functions on  $\{\beta\}$ .

In order to define a k-system, we do the following transformation. We define

$$k(\alpha) = f(\alpha)/\tau, \text{ for all } \alpha, \quad (1.11a)$$

so that

$$0 \leq k(\alpha) \leq 1, \text{ for all } \alpha, \text{ and} \quad (1.11b)$$

$$\sum_{\alpha} k(\alpha) = 1. \quad (1.11c)$$

Now a k-system is defined as the following tuple:

$$(\tau, \{v_i\}, \{\alpha\}, \{\beta\}, k, \{^m k\}), \quad (1.12)$$

where  $\tau$  is a transformation factor,  $\{v_i\}$  is a set of variables,  $\{\alpha\}$  is a set of states,  $\{\beta\}$  is a set of substates,  $k$  is a function on  $\{\alpha\}$ , and  $\{^m k\}$  are functions on  $\{\beta\}$ .

From the above description, it is obvious that a k-system is a dimensionless system, though it is isomorphic to a g-system. There is no loss or gain of information in switching from one system to the other [38]. The goal of k-systems is to utilize the framework of probabilistic RA in order to develop solution procedures for g-systems.

#### 1.4 Why RA and Why K-systems

RA deals primarily with systems and subsystems [9], more specifically, with states and substates [33-41] and the interrelationships among themselves. The state/substate paradigm of RA has led to a departure from the classical statistical approach, has initiated the design of more powerful algorithms, and has thus provided new insights into the structure and dynamics of systems. Contrary to RA, classical statistical approaches emphasize variables [40]. Concepts of state and substate in RA have no equivalents in the classical framework. For instance, in RA, an event is the

occurrence of a state (or substate), but an event is not a state by itself. A state will exist regardless of its occurrence, and states can be referred to without implying their occurrence. Thus, RA has the concept of independent events which has a classical equivalent, and the concept of independent states which has no classical equivalent. The concept of independent states has played a very significant role in the design of algorithms and procedures in RA. A state viewed apart from its occurrence is important everywhere in RA, illustratively, in null extensions [33-41] or in evaluating the cognitive contents of system substates [37]. Maximum entropy algorithms designed by Jones [31-32] have been employed successfully only after the advent of RA.

The inherent advantage of the RA methodology, more specifically of k-systems, over the classical techniques is that it does not assume any structure in the data. RA is as good as the data it considers. On the contrary, classical techniques induce some kind of model on the data and hence introduce extraneous information. For instance, analysis of variance, one of the most powerful techniques in the field of statistical inference, assumes a linear model on the data under consideration [40].

### **1.5 Expert Systems and Knowledge Acquisition**

Expert systems are special purpose computer programs that use specialized symbolic reasoning to solve difficult problems effectively [24, 30, 51]. In contrast with conventional general purpose computer programs, expert systems are characterized by (a) use of specialized knowledge about a particular application domain, such as medicine, (b) use of symbolic reasoning (rather than numerical computations), and (c) performance more competent than that of non-experts. Expert systems deal mostly with qualitative and nominal data rather than quantitative data. Unlike traditional computer programs that employ algorithms, expert systems use heuristic reasoning. They attack complex problems that are very difficult to solve perfectly, and provide good, though not necessarily optimal, answers to those problems.



Learning is the ability of a system, biological or physical, to change itself to improve its problem solving capability. Improvement of a system's problem solving capability includes, but is not limited to, performing generalization, providing better quality results, or doing the same thing faster. Machine learning is concerned with the study of computational aspects of the learning process, and to build machine or computer programs that are capable of learning via computation. Learning is particularly important in the context of expert systems. Given a database of a certain application domain, an algorithm  $A$  can be employed to directly learn the rules from the database. This simplifies the problem of the knowledge acquisition bottleneck in acquiring the knowledge directly from the expert(s). Given a knowledge base (domain theory), another algorithm  $B$  can be employed to revise the domain theory, if necessary [64].

Let  $A$  be a learning algorithm.  $A$  is given a set of positive and negative instances of a given concept.  $A$  is supposed to find a hypothesis in the hypothesis space that best describes the concept. Let  $v$  be a positive instance in the instance space. Let  $A$  be a statistical (probabilistic) learning algorithm. Then,  $A$  attempts to find a probability distribution, or a probabilistic mapping  $F$ , between the space of instances and space of hypotheses, so that  $\text{prob}(F(v) = 1) \geq 1 - \delta$ . Here  $\delta$  is a very small number, close to zero, and we refer to it as Baye's misclassification rate [18, 64]. On the other hand, if  $A$  is a symbolic learning algorithm, it will attempt to find a deterministic mapping, such as a boolean function  $F$ , between the two spaces, so that  $F(v) = 1$  for all  $v$ . Obviously, the symbolic learning algorithm  $A$  implicitly assumes that  $\delta$ , the Baye's misclassification rate is zero. In simple words, this translates to the assumption that given a complete set of instances, it is always possible to perfectly classify each attribute in terms of the other attributes, which is not necessarily true [64].

## CHAPTER 2

### SOME NEW RESULTS IN RECONSTRUCTABILITY ANALYSIS

#### 2.1 Determination of Reconstruction Families Using independent Information

One of the problems in the reconstructability analysis is the identification problem that requires the determination of reconstruction families. Jones [33] provides an innovative method of computing reconstruction families using only independent information. This method gives a more efficient form of matrix equations for the reconstruction hypothesis. The resultant matrix is in a triangular form. In this section, we first outline the above method, and then provide a generalization to this method in the sense that the choice of substates is not limited by a particular ordered set  $P_\phi$ . Any substate belonging to the same equivalence class as a member of  $P_\phi$  can be employed in the process of reconstruction.

##### 2.1.1 Introduction

Let  $B$  be an overall system, and let  $f$  be a probabilistic behavior function. Let  $S = \{ {}^k B \}$  be a reconstruction hypothesis of  $B$ . Let  $A$  be the set of aggregate states of  $B$  and  ${}^k A$  be the set of substates of  ${}^k B$ . Then  ${}^k f$  must satisfy the following conditions:

$${}^k f(\beta) = \sum_{\alpha > \beta} f(\alpha), \quad (2.1a)$$

$$\sum_{\alpha \in A} f(\alpha) = 1, \text{ and} \quad (2.1b)$$

$$f(\alpha) \geq 0 \text{ for all } \alpha \in A, \quad (2.1c)$$

where  $\beta \in {}^k A, \alpha \in A$ , and  $\alpha > \beta$  implies that  $\alpha$  is a superstate of  $\beta$  or  $\beta$  is a substate of  $\alpha$ .

The above two conditions describe linear algebraic equations whose variables are the probabilities of aggregate states of  $B$ . Each solution to the set of these equations uniquely represents a member of reconstruction family  $B_S$  of  $S$ . Now with regard to the structure system  $\{^k B\}$ , we define the following terms [33]:

**Definition 2.1** Let power set  $P_V$  for  $\{^k V\}$  be the set of all subsets of  $^k V$ .

**Definition 2.2** Let  $P_f$  for  $\{^k V\}$  be the set of behavior functions corresponding to the elements of the power set of  $\{^k V\}$ .

**Definition 2.3** Let  $^r f$  and  $^s f$  be in  $P_f$  then  $r < s$  if and only if  $|^r V| \leq |^s V|$ .

**Definition 2.4** Given any set of variables, let  $\phi$ -states of this set be defined as the states with all non-zero variable values.

**Definition 2.5** The set  $P_\phi$  is the set of probabilities associated with the  $\phi$ -states corresponding to the power set for  $\{^k V\}$ .

**Definition 2.6** Let  $\beta \in ^k A$  be a substate of  $\alpha \in A$ . Let  $v_i \in V$  and  $v_i \notin ^k V$ . Then  $\alpha$  is a null extension of  $\beta$  if and only if  $v_i = 0$  for all applicable  $i$ 's.

**Definition 2.7** Two substates are said to be equivalent if and only if they have same null extension.

Consider the example given in section 1.2. For  $^{23}V = \{v_2, v_3\}$  and  $^{12}V = \{v_1, v_2\}$ ,  $P_f = \{^1 f(v_1), ^2 f(v_2), ^3 f(v_3), ^{12} f = ^4 f(v_1, v_2), ^{23} f = ^5 f(v_2, v_3), ^{13} f = ^6 f(v_1, v_3)\}$ .

This is because we induce a partial ordering on the members of  $P_f$ . We know that each  $v_i$  must take a finite number of values. Without loss of generality, we assume that each  $v_i$  takes values from the set  $\{0, 1, 2, \dots, n_i\}$ . If, in the set  $\{v_1, v_2\}$ ,  $v_1$  and  $v_2$  take values 0, 1, 2, then the corresponding  $\phi$ -states are 11, 12, 21, 22. Clearly these  $\phi$ -states are ordered using the very last strategy. First, we vary the second variable and then the first variable. Finally, we get

$$P_\phi = \{^1 f(1), ^2 f(1), ^3 f(1), ^3 f(2), ^4 f(11), ^5 f(11), ^5 f(12), ^6 f(11), ^6 f(12)\}.$$

### 2.1.2 Determination of Triangular Matrix

Now we put the above information in the form  $MQ = W$ , where  $M$  is a coefficient matrix whose entries are zeroes and ones,  $Q$  is the column vector of unknown  $f(\alpha)$ 's and  $W$  is a column vector of the elements of  $P_\phi$  preceded by 1. Assume there are  $m$  elements in  $P_\phi$ . Following is an algorithm by Jones [33] to determine the reconstruction family in the form of linearly independent matrix equations.

#### 2.1.2.1 Jones' Algorithm

- [1] Let  $W_1 = 1$ ,  $W_2 =$  first element of  $P_\phi$ ,  $W_3 =$  second element of  $P_\phi$ , and so on .
- [2] Let  $Q_1 = f(000 \dots 0)$ . Let  $\alpha$  be the null extension of the substate associated with  $W_i$ . Then  $Q_i = f(\alpha)$ , for  $3 \leq i \leq m + 1$ .
- [3] If any  $f(\alpha)$ 's are left, arrange them in any order.
- [4] Fill in the matrix using equation (2.1a).

We observe that  $M$  will always be an upper triangular matrix. The matrix equations in the form  $MQ = W$  describe ten linearly independent equations with 12 variables. These equations can easily be solved and they allow two degrees of freedom.

$$W = \begin{bmatrix} 1 \\ {}^1f(1) \\ {}^2f(1) \\ {}^3f(1) \\ {}^3f(2) \\ {}^4f(11) \\ {}^5f(11) \\ {}^5f(12) \\ {}^6f(11) \\ {}^6f(12) \end{bmatrix} = \begin{bmatrix} 1 \\ 0.57 \\ 0.55 \\ 0.34 \\ 0.35 \\ 0.37 \\ 0.18 \\ 0.23 \\ 0.20 \\ 0.17 \end{bmatrix}; \quad Q = \begin{bmatrix} f(000) \\ f(100) \\ f(010) \\ f(001) \\ f(002) \\ f(110) \\ f(110)^{011} \\ f(011)^{012} \\ f(101) \\ f(102) \\ f(111) \\ f(112) \end{bmatrix}.$$

The matrix equations are therefore

$$\begin{bmatrix}
 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\
 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1
 \end{bmatrix} \times \begin{bmatrix} f(000) \\ f(100) \\ f(010) \\ f(001) \\ f(002) \\ f(110) \\ f(011) \\ f(012) \\ f(101) \\ f(102) \\ f(111) \\ f(112) \end{bmatrix} = \begin{bmatrix} 1 \\ 0.57 \\ 0.55 \\ 0.34 \\ 0.35 \\ 0.37 \\ 0.18 \\ 0.23 \\ 0.20 \\ 0.17 \end{bmatrix}.$$

All the solutions of these equations form the reconstruction family of  $B_S$ . Use of the constraint  $f(\alpha) \geq 0$  will yield half space constraints on these two free variables.

### 2.1.2.2 A New Algorithm

Now we give an algorithm which accomplishes the same goal. However, the new algorithm constructs the set  $W$  differently. The concept of the null extension is used to divide the whole state space into disjoint equivalence classes as follows.

Equivalence Class	Null Extension	Equivalence Class	Null Extension
$\{ {}^1(0), {}^2(0), {}^3(0), {}^4(00), {}^5(00), {}^6(00) \}$	000	$\{ {}^5(12) \}$	012
$\{ {}^3(1), {}^5(01), {}^6(01) \}$	001	$\{ {}^6(11) \}$	101
$\{ {}^3(2), {}^6(02), {}^5(02) \}$	002	$\{ {}^6(12) \}$	102
$\{ {}^2(1), {}^4(01), {}^5(10) \}$	010	$\{ {}^4(11) \}$	110
$\{ {}^1(1), {}^5(10), {}^6(10) \}$	100	$\{ \}$	112
$\{ {}^5(11) \}$	011	$\{ \}$	111

It is clear that each member of  $P_\phi$  belongs to a different equivalence class. That is, no two members of  $P_\phi$  belong to the same class. Now, we construct  $W$ . The first element of  $W$  is equal to 1, which relates to relation (2.1b). Let  $W_i$  be any element belonging to the equivalence class associated with the  $(i-1)$ th element of  $P_\phi$ . It is worth mentioning that we need not be strict in our choice of states. Rather we can pick any state which is equivalent to an element in  $P_\phi$ .  $W$  is computed as follows,

$$W = \begin{bmatrix} 1 \\ {}^4f(10) \\ {}^4f(01) \\ {}^5f(01) \\ {}^5f(02) \\ {}^6f(11) \\ {}^4f(11) \\ {}^5f(11) \\ {}^5f(12) \\ {}^6f(12) \end{bmatrix} = \begin{bmatrix} 1 \\ 0.20 \\ 0.18 \\ 0.16 \\ 0.12 \\ 0.37 \\ 0.18 \\ 0.23 \\ 0.20 \\ 0.17 \end{bmatrix}.$$

Obviously, there will not be any change in vector  $Q$ , though matrix  $M$  may be different. Following is the matrix equation

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} f(000) \\ f(100) \\ f(010) \\ f(001) \\ f(002) \\ f(110) \\ f(011) \\ f(012) \\ f(101) \\ f(102) \\ f(111) \\ f(112) \end{bmatrix} = \begin{bmatrix} 1 \\ 0.20 \\ 0.18 \\ 0.16 \\ 0.12 \\ 0.37 \\ 0.18 \\ 0.23 \\ 0.20 \\ 0.17 \end{bmatrix}.$$

We again have an upper triangular matrix  $M$  giving ten linearly independent equations in twelve unknowns, implying two free variables. Using the constraint  $f(\alpha) \geq 0$  yields half space constraints on two free variables. We assert that the new algorithm is more general as it does not limit the choice of states  $P_\phi$  but allows us to pick any substate from the equivalence classes corresponding to the elements in  $P_\phi$ .

### 2.1.3 Concluding Remarks

Our algorithm is more general than one proposed by Jones [33]. Use of higher order states makes the matrix sparser and thus could improve upon the computation, enabling us to reduce the space and time required to determine the reconstruction family, especially when there are many  $v_i$ 's included.

Now, the question is whether or not we lose any information by choosing equivalents of  $\phi$ -states instead of  $\phi$ -states. As proven in [33], there is no loss of information in constructing the set  $P_\phi$ . It is observed that the number of  $\phi$ -states associated with the power set for  ${}^iV$  is equal to the number of unknowns minus one minus the number of unknowns with all non-zero valued variables (not the number of unknowns minus one as mentioned by Jones [33]). Let us consider aggregate states for  ${}^iV = \{v_1, v_2, \dots, v_n\}$  which have a zero value for one or more variables but not for all. This aggregate state can be placed into one to one correspondence with a  $\phi$ -state for which all zero-valued variables are deleted. Alternatively, this aggregate state can be placed into one to one correspondence with an equivalent of a  $\phi$ -state for which a subset of zero-valued variables is deleted. Note that an equivalent of a  $\phi$ -state will consist of more variables than  $\phi$ -states. This establishes a one to one correspondence between all aggregate states of  ${}^iV$  ( except  $(00 \dots 0)$  and those with no zero-valued variables ) and the  $\phi$ -states associated with  $P_\phi$ . Thus there is no loss of information in forming  $P_\phi$ . The solution is achieved in the form of a possibly different set of linearly independent equations.

## 2.2 An Algorithm for Selection Functions

In this section we give a reconstruction algorithm based on [9], where the system function is a selection function, i.e., only the occurrence or the non-occurrence of a system state is recorded. Klir [9] provides two algorithms namely, algorithm 1 and algorithm 2, to compute the unbiased reconstruction. Algorithm 1 is based on Ashby's inverse procedure [2]. Algorithm 2 uses a relational join in order to compute the overall maximum entropy reconstruction.

### 2.2.1 Introduction

Let  $B = (V, W, s, A, Q, f)$  be a selection behavior system. Let  $S = \{ {}^k B \} = \{ ({}^k V, {}^k W, {}^k s, {}^k A, {}^k Q, {}^k f) \mid k \in \{1, 2, \dots, q\} \}$  be a reconstruction hypothesis for  $B$ . By the definition of a selection function, functions  $f$  and  $\{ {}^k f \}$  can take only 0 or 1 values. 0 indicates the occurrence of a state and 1 indicates the non-occurrence of a state in the system or subsystem. First we outline the algorithms given by Klir [9]. Then we shall propose a new algorithm to solve this problem.

### 2.2.2 Algorithm 1

Let  $S$  be a reconstruction hypothesis for a selection behavior system. Before we describe algorithm 1, we define the following terms which will be used in the algorithm:

$$F = \{ \alpha \mid f(\alpha) = 1 \}, \quad (2.2a)$$

$$F_c = A - F = \{ \alpha \mid f(\alpha) = 0 \}, \quad (2.2b)$$

$${}^k F = \{ \beta \mid {}^k f(\beta) = 1 \}, \quad (2.2c)$$

and

$${}^k F_c = {}^k A - {}^k F = \{ \beta \mid {}^k f(\beta) = 0 \}. \quad (2.2d)$$



Now, the unbiased reconstruction is determined as follows

- [1]  $X := A; k := 1;$
- [2] for all  $\alpha \in X$  and for all  $\beta \in {}^k F_c$ , if  $\beta < \alpha$  then  $X := X - \{\alpha\};$
- [3] if  $k < N_q$  then  $k := k + 1;$  go to step [2]
- [4] stop;  $F_S := X;$   $F_S$  is the unbiased reconstruction.

### 2.2.3 Algorithm 2

Let  $S$  be a reconstruction hypothesis for a selection behavior system. Before we describe algorithm 2, we define some additional terms which will be used in this algorithm. Let  $R_1 \subset X_1 \times X_2$  and  $R_2 \subset X_2 \times X_3$  be two binary relations. Then their join  $R_1 * R_2$  is a ternary relation defined as

$$R_1 * R_2 = \{(\alpha, \beta, \gamma) \mid (\alpha, \beta) \in R_1, (\beta, \gamma) \in R_2\}, \quad (2.3a)$$

$$R_1 * R_2 = \{(\beta, \gamma) \mid \beta \in R_1, (\beta, \gamma) \in R_2\} \text{ when } X_1 = \Phi \text{ and } R_1 \subset X_2, R_2 \subset X_2 \times X_3, \quad (2.3b)$$

$$R_1 * R_2 = \{(\alpha, \gamma) \mid \alpha \in R_1, \gamma \in R_2\} \text{ when } X_2 = \Phi \text{ and } R_1 \subset X_1, R_2 \subset X_3. \quad (2.3c)$$

The unbiased reconstruction  $F_S$  is determined as follows:

- [1]  $k := 2; R := {}^1 F;$
- [2] convert  ${}^k F$  and  $R$  into appropriate binary relations and perform  $R := {}^k R * R;$
- [3] if  $k < N_q$  then  $k := k + 1;$  go to step [2];
- [4] stop;  $F_S := R;$   $F_S$  is the unbiased reconstruction.

### 2.2.4 Algorithm 3

Given a reconstruction hypothesis  $S$  with behavior functions  $\{{}^k f\}, k \in N_q$ , determine the unbiased reconstruction  $f$  implied by  $S$ . Following is our algorithm to compute the unbiased reconstruction.

- [1]  $f(\alpha) := 1$  for all  $\alpha$  in the system,  $\alpha$  being an aggregate state;  $k := 1$ ;
- [2] choose  ${}^k f$ ; For all aggregate states  $\alpha$  and for all  $\beta$  related to  ${}^k f$  such that  $\beta < \alpha$  do the following:  $f(\alpha) := \min(f(\alpha), f(\beta))$ ;
- [3] if  $k < N_q$  then  $k := k + 1$ ; go to step [2];
- [4] stop;  $f_S := f$ ;  $f_S$  is the unbiased reconstruction.

Now we must prove that the above algorithm indeed determines the unbiased reconstruction. It will suffice to show that the result obtained by Algorithm 3 is essentially the same as one obtained by using Ashby's procedure. In order to show it we first define the following.

**Definition 2.8** An extension of  ${}^k F_c$  with respect to variables in  $V - {}^k V$  consists of all overall states whose projections are the states in  ${}^k F_c$ . This is denoted by  $[{}^k F_c \uparrow V - {}^k V]$ .

Let  $X = \{\alpha \mid f(\alpha) = 0 \text{ after performing } f * {}^1 f\}$ . In order to show that the result of Algorithm 3 is same as Ashby's procedure, we prove the following inequality

$$[{}^k F_c \uparrow V - {}^k V] = X. \quad (2.4)$$

Let  $\alpha \in [{}^k F_c \uparrow V - {}^k V] \Rightarrow \alpha \in A, (\text{there exists a } \beta \in {}^k F_c)(\alpha > \beta)$  (by definition of an extension)  $\Rightarrow f(\alpha) = 0 \Rightarrow f(\alpha) \in X$ . Therefore  $[{}^k F_c \uparrow V - {}^k V] \subseteq X$ . Let  $\alpha \in X \Rightarrow f(\alpha) = 0 \Rightarrow$  As per Ashby's procedure, there exists a  $\beta \in {}^k F_c$  such that  $\alpha > \beta \Rightarrow \alpha \in [{}^k F_c \uparrow V - {}^k V] \Rightarrow X \subseteq [{}^k F_c \uparrow V - {}^k V]$ . Thus  $\Rightarrow X = [{}^k F_c \uparrow V - {}^k V]$ . We can make same arguments for joins of other  $\{{}^k f\}$ . This proves the following theorem.

**Theorem 2.1** Given a consistent reconstruction hypothesis  $S$  with behavior functions  $\{{}^k f\}, k \in N_q$ , Algorithm 3 computes the unbiased reconstruction  $f_S$  implied by  $S$ .

Following example displays how Algorithm 3 works. Let  $S = \{({}^kV, {}^kW, {}^kS, {}^kA, {}^kQ, {}^kf)\}$  be a structure system such that  ${}^{12}V = \{v_1, v_2\}$ ,  ${}^{23}V = \{v_2, v_3\}$ ,  ${}^{13}V = \{v_1, v_3\}$ ,  $s(v_1) = \{0, 1\}$ ,  $s(v_2) = \{0, 1\}$ , and  $s(v_3) = \{0, 1, 2\}$ .  $\{ {}^kf \}$  are given below.

$v_1$	$v_2$	${}^{12}f$	$v_2$	$v_3$	${}^{23}f$	$v_1$	$v_3$	${}^{13}f$
—	—	—	—	—	—	—	—	—
0	0	1	0	0	1	0	0	1
0	1	1	0	1	1	0	1	1
1	0	1	0	2	1	0	2	1
1	1	0	1	0	1	1	0	1
			1	1	1	1	1	0
			1	2	0	1	2	0

Applying the above algorithm on the given example yields

$v_1$	$v_2$	$v_3$	$f$				$v_1$	$v_2$	$v_3$	$f$
—	—	—	—				—	—	—	—
0	0	0	1				0	0	0	1
0	0	1	1				0	0	1	1
0	0	2	1	$v_1$	$v_2$	${}^{12}f$	0	0	2	1
0	1	0	1	—	—	—	0	1	0	1
0	1	1	1	*	0	0	1	*	0	1
0	1	2	1		0	1	1		0	2
1	0	0	1		1	0	1		1	0
1	0	1	1		1	1	0		1	1
1	0	2	1				1		1	2
1	1	0	1				1		1	0
1	1	1	1				1		1	1
1	1	2	1				1		1	2

$$\begin{array}{ccccccc}
& v_1 & v_2 & v_3 & f & & \\
& \hline & \hline & \hline & \hline & & & \\
& 0 & 0 & 0 & 1 & & & \\
& 0 & 0 & 1 & 1 & v_1 & v_3 & {}^{13}f \\
& 0 & 0 & 2 & 1 & \hline & \hline & \hline \\
& 0 & 1 & 0 & 1 & 0 & 0 & 1 \\
& 0 & 1 & 1 & 1 & 0 & 1 & 1 \\
= & 0 & 1 & 2 & 0 & * & 0 & 2 & 1 \\
& 1 & 0 & 0 & 1 & 1 & 0 & 1 \\
& 1 & 0 & 1 & 1 & 1 & 1 & 0 \\
& 1 & 0 & 2 & 1 & 1 & 2 & 0 \\
& 1 & 1 & 0 & 0 & & & \\
& 1 & 1 & 1 & 0 & & & \\
& 1 & 1 & 2 & 0 & & & 
\end{array}
=
\begin{array}{cccc}
v_1 & v_2 & v_3 & f \\
\hline & \hline & \hline & \hline \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 \\
0 & 0 & 2 & 1 \\
0 & 1 & 0 & 1 \\
0 & 1 & 1 & 1 \\
0 & 1 & 2 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 \\
1 & 0 & 2 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 \\
1 & 1 & 2 & 0
\end{array}
.$$

### 2.2.5 Concluding Remarks

Basically, all the three algorithms are the same and will converge at the same rate. No iterative procedure is required to compute the unbiased reconstruction. The algorithm we propose is easy to understand and simple to use because it starts with a constant distribution. Moreover, this algorithm is based on certain possibilistic measures rather than a simple relational join. Therefore, it is possible to compute the unbiased reconstruction when the information is not given in its entirety. We shall dwell upon these issue in chapter 3.

### 2.3 Reconstruction of Probabilistic Behavior Functions

In this section we give a reconstruction algorithm based on [9] and [33], where the system function is a probabilistic behavior function. Probabilities of occurrence of system states are recorded. Cavallo and Klir [9] provide an algorithm for computing unbiased reconstruction from a given reconstruction hypothesis. They use a probabilistic version of Ashby's join procedure.

### 2.3.1 Introduction

Let  $B = (V, W, s, A, Q, f)$  be a probabilistic behavior system as defined in (1.1). Let  $S = \{ {}^k B \} = \{ ({}^k V, {}^k W, {}^k s, {}^k A, {}^k Q, {}^k f) \mid k \in \{1, 2, \dots, q\} \}$  be a reconstruction hypothesis for  $B$ . By the definition of a probabilistic behavior function, function  $f$  and  $\{ {}^k f \}$  record the probability of occurrence of states and the substates, respectively, and sum to one. First, we outline the algorithm given Cavallo and Klir [9]; then we give a new algorithm, which is simple to use.

Given two probabilistic behavior functions  ${}^1 f: X_1 \times X_2 \rightarrow [0, 1]$  and  ${}^2 f: X_2 \times X_3 \rightarrow [0, 1]$ , their join  ${}^1 f * {}^2 f: X_1 \times X_2 \times X_3 \rightarrow [0, 1]$  is defined as

$$[{}^1 f * {}^2 f](\alpha, \beta, \gamma) = {}^1 f(\alpha, \beta) \cdot {}^2 f(\gamma \mid \beta), \quad (2.5a)$$

$$[{}^1 f * {}^2 f](\beta, \gamma) = {}^1 f(\beta) \cdot {}^2 f(\gamma \mid \beta) \text{ if } X_1 = \Phi \text{ and } {}^1 f: X_2 \rightarrow [0, 1], \quad (2.5b)$$

$$[{}^1 f * {}^2 f](\alpha, \gamma) = {}^1 f(\alpha) \cdot {}^2 f(\gamma) \text{ if } X_2 = \Phi, {}^1 f: X_1 \rightarrow [0, 1] \text{ and } {}^2 f: X_3 \rightarrow [0, 1]. \quad (2.5c)$$

Here  $\alpha, \beta, \gamma$  denotes the concatenation of disjoint substates  $\alpha, \beta, \gamma$ . Note that  ${}^2 f(\gamma \mid \beta)$  stands for the conditional probability of  $\gamma$  given  $\beta$ . That is,

$${}^2 f(\gamma \mid \beta) = {}^2 f(\beta, \gamma) / \sum_w {}^2 f(\beta, w). \quad (2.5d)$$

Cavallo and Klir [9] give two algorithms to compute unbiased reconstruction. The first, called a basic join procedure, computes the unbiased reconstruction and checks for conformity with  $S$ . If  $f$  does not conform to  $S$ , then it employs the second algorithm, called an iterative join procedure [9].

### 2.3.2 Basic Join Procedure

Let  $S$  be a locally consistent reconstruction hypothesis with probabilistic behavior functions  $\{ {}^k f \} (k \in N_q)$ . Then the unbiased reconstruction  $f_S$  implied by  $S$  is determined as follows [9]:

- [1]  $k := 2; f := {}^1 f$  ;
- [2] adjust arguments of  ${}^k f$  and  $f$  ;  $f := {}^k f * f$  ;
- [3] if  $k < N_q$  then  $k := k + 1$ ; go to step [2];
- [4] stop; if  $[f \downarrow {}^k V] = {}^k f$  for all  $k \in N_q$  then  $f_S := f$ ;  $f_S$  is the unbiased reconstruction  
else  $f$  does not conform to  $S$ ; employ iterative joint procedure.

### 2.3.3 Iterative Join Procedure

Let  $S$  be a locally consistent reconstruction hypothesis with probabilistic behavior functions  $\{ {}^k f \} (k \in N_q)$ . Let  $\delta \in [0, 1]$  be a small number. Let  $f$  be the result of the basic join procedure. The iterative join procedure for computing the unbiased reconstruction is as follows:

- [1]  $k := 1; i := 1; f_0 = f$ ;
- [2] adjust arguments of  ${}^k f$  and  $f_{i-1}$  ;  $f_i := {}^k f * f_{i-1}$ ;
- [3] if  $i \bmod q \neq 0$  then  $i := i + 1; k := k + 1$ ; go to step [2];
- [4] if  $|f_i(\alpha) - f_{i-q}(\alpha)| \geq \delta$  for some  $\alpha$  then  $i := i + 1; k := (k + 1) \bmod q + 1$ ; go to step [2];
- [5] stop; if  $\sum_i f_i(\alpha) = 1$  then  $f_i(\alpha) = f_S(\alpha) \pm \delta$  for each  $\alpha$  where  $f_S$  is the unbiased reconstruction implied by  $S$ , otherwise  $S$  is globally inconsistent and has no reconstruction.

### 2.3.4 Unified Join Procedure

We will give a unified procedure which exhibits the features of both the basic and the iterative join procedures. We discover that the basic join procedure is not at all necessary, as we have incorporated it into the unified join procedure. If there are no loops in the reconstruction hypothesis, the unified join procedure will compute the unbiased reconstruction after the first iteration and will automatically stop. If the

reconstruction hypothesis consists of loops, then the procedure will continue until convergence is achieved and the unbiased reconstruction is computed. The unified join procedure is:

- [1]  $k := 1; i := 1; f_0(\alpha) := 1/n$  where  $n$  is the number of aggregate states in the system;
- [2]  $f_i(\alpha) = f_{i-1}(\alpha) *^k f$  for all  $\alpha$ .
- [3] if  $|f_i(\alpha) - f_{i-1}(\alpha)| \geq \delta$  for some  $\alpha$  then  $i := i + 1; k = (k + 1) \bmod q + 1$ ; go to step [2];
- [4] stop; if  $\sum_i f_i(\alpha) = 1$  then  $f_i(\alpha) = f_S(\alpha) \pm \delta$  for each  $\alpha$ ,  $f_S$  is unbiased reconstruction; otherwise  $S$  is globally inconsistent and no reconstruction exists for  $S$ .

Note that unified join procedure is simpler to use. It requires only the computation of a degenerate form of conditional probabilities. That is,  $f_i(\alpha) = f_{i-1}(\alpha) *^k f$  can be written as

$$f_i(\alpha) = \frac{f_{i-1}(\alpha)}{[f_{i-1} \downarrow^k V](^k \alpha)} \cdot ^k f(^k \alpha), \quad (2.6a)$$

$$\text{or } f_i(\alpha) = \frac{f_{i-1}(\alpha)}{\sum_{^k \alpha} f(\alpha)} \cdot ^k f(^k \alpha), \quad (2.6b)$$

$$\text{or } f_i(\alpha) = f_{i-1}(\alpha \mid ^k \alpha) \cdot ^k f(^k \alpha). \quad (2.6c)$$

In order to prove our result, we observe that (2.6a) is simply an iterative scheme as proposed by Brown [3, 6]. As per Brown's findings, this scheme converges to a behavior function which conforms to the reconstruction hypothesis  $S$  and is the unbiased reconstruction from  $S$ . Thus, we conclude the following theorem.

**Theorem 2.2** Let  $S$  be a consistent reconstruction hypothesis. Then the unified join procedure computes the unbiased reconstruction  $f_S$  implied by  $S$ .

Consider an example applying unified join procedure on a structure system involving variables  $v_1, v_2$  and  $v_3$ . The subsystem functions  $\{^k f\}$  are given below.

$v_1$	$v_2$	$^{12}f$	$v_2$	$v_3$	$^{23}f$
$\overline{\quad}$	$\overline{\quad}$	$\overline{\quad}$	$\overline{\quad}$	$\overline{\quad}$	$\overline{\quad}$
0	0	0.1	0	0	0.2
0	1	0.0	0	1	0.0
0	2	0.0	1	0	0.1
0	3	0.3	1	1	0.1
1	0	0.1	2	0	0.0
1	1	0.2	2	1	0.0
1	2	0.0	3	0	0.4
1	3	0.3	3	1	0.2

Applying our algorithm on this example yields

$v_1$	$v_2$	$v_3$	$f$			
$\overline{\quad}$	$\overline{\quad}$	$\overline{\quad}$	$\overline{\quad}$			
0	0	0	0.0625			
0	0	1	0.0625			
0	1	0	0.0625	$v_1$	$v_2$	$^{12}f$
0	1	1	0.0625	$\overline{\quad}$	$\overline{\quad}$	$\overline{\quad}$
0	2	0	0.0625	0	0	0.1
0	2	1	0.0625	0	1	0.0
0	3	0	0.0625	0	2	0.0
0	3	1	0.0625	0	3	0.3
1	0	0	0.0625	1	0	0.1
1	0	1	0.0625	1	1	0.2
1	1	0	0.0625	1	2	0.0
1	1	1	0.0625	1	3	0.3
1	2	0	0.0625			
1	2	1	0.0625			
1	3	0	0.0625			
1	3	1	0.0625			



	$v_1$	$v_2$	$v_3$	$f$		$v_1$	$v_2$	$v_3$	$f$
	0	0	0	0.5		0	0	0	0.1
	0	0	1	0.5		0	0	1	0.0
	0	1	0	0.0	$v_2$	0	1	0	0.0
	0	1	1	0.0	$v_3$	0	1	1	0.0
	0	2	0	0.0	$^{23}f$	0	2	0	0.0
	0	2	1	0.0	—	0	2	1	0.0
	0	3	0	0.5	0	0	3	0	0.2
=	0	3	1	0.5	0	0	3	1	0.0
	1	0	0	0.5	0	1	0	0	0.1
	1	0	1	0.5	0	1	1	0	0.0
	1	1	0	1.0	0	1	2	0	0.1
	1	1	1	1.0	0	1	3	0	0.0
	1	2	0	0.0	0	1	2	0	0.0
	1	2	1	0.0	0	1	2	1	0.0
	1	3	0	0.5	0	1	3	0	0.2
	1	3	1	0.5	0	1	3	1	0.1

The computation of conditional probabilities is implicit. In this case the algorithm terminates one iteration after it performs joints with every  $^k f$ . In the cases where there are loops in the structure systems, the computation will terminate after the desired accuracy is achieved in the process of computing the unbiased reconstruction.

### 2.3.5 Concluding Remarks

The inception of the unified join procedure renders the computation of unbiased reconstruction trivial. No explicit computation of conditional probabilities is required. Further, in contrast to the iterative join procedure, the algorithm terminates as soon as the unbiased reconstruction is achieved. For example, suppose in some situation the unified join procedure terminates after  $f_0 * ^1 f * ^2 f * ^3 f * ^1 f * ^2 f * ^3 f * ^1 f$ . For the same case, the iterative join procedure will terminate after  $f * ^1 f * ^2 f * ^3 f * ^1 f * ^2 f * ^3 f * ^1 f * ^2 f * ^3 f$  with the same result, causing

unnecessary computation. We conclude with the remark that the unified join procedure provided here is the unified version of the basic join procedure and the iterative join procedure. One procedure will do for all kinds of probabilistic functions. There is no need to employ two separate procedures.

## 2.4 Reconstructability Analysis for U-structures

In this section, we define a special class of structure systems, called U-structures, and give an algorithm to compute unbiased reconstruction for such structures.

**Definition 2.9** Let  $S = \{ {}^k B \} = \{ ({}^k V, {}^k W, {}^k s, {}^k A, {}^k Q, {}^k f) \mid k \in \{1, 2, \dots, q\} \}$  be a structure system. Subsystems  ${}^i B$  and  ${}^j B$  are said to be coupled together by the variables in the set  ${}^i V \cap {}^j V$  if  ${}^i V \cap {}^j V$  is non-empty, otherwise  ${}^i B$  and  ${}^j B$  are said to be uncoupled.

**Definition 2.10** Let  $S = \{ {}^k B \}$  be a structure system.  $S$  is said to be uncoupled structure (or  $U$ -structure) if and only if  $\bigcap_{k \in N_q} {}^k V = \Phi$ .

We shall now propose a reconstruction algorithm for U-structures when only independent information is employed. We name this algorithm the U-algorithm.

### 2.4.1 U-algorithm

Let  $S = \{ {}^k B \}$  be an  $U$ -structure. We compute the unbiased reconstruction implied by  $S$  using only independent information  $\{ {}^k B \}$ . The algorithm proceeds as follows:

- [1]  $f(\alpha) := 1$  for all  $\alpha$ ;
- [2] for each  $\alpha$  and  ${}^k \alpha$  in  $\{ {}^k \alpha \}$  if  $\alpha > {}^k \alpha$  then  $f(\alpha) = f(\alpha) \cdot {}^k f({}^k \alpha)$ ;
- [3] for each  $\alpha$  with one or more zero and for each 0 in  $\alpha$  if  $v_i$  in  $\alpha$  is 0 then  $f(\alpha) = f(\alpha) \cdot {}^i f_c({}^i \alpha)$ ;

[4] stop;  $f_S := f$ ;  $f_S$  is the unbiased reconstruction.

Notice that, in the above algorithm

$${}^i f_c({}^i \alpha) = 1 - \sum_{{}^i \beta \neq 0} {}^i f({}^i \alpha). \quad (2.7)$$

In fact, the algorithm describes a *partial join procedure*. The algorithm terminates after computing the unbiased reconstruction, and no iterative join is required.

Consider the following example.

$v_1$	${}^1 f$	$v_2$	${}^2 f$
—	—	—	—
0	0.3	0	0.1
1	0.2	1	0.3
2	0.5	2	0.6

Our algorithm works on this example as follows:

	$v_1$	$v_2$	$f$		$v_1$	$v_2$	$f$		$v_1$	$v_2$	$f$
	—	—	—		—	—	—		—	—	—
	0	0	1		0	0	1		0	0	0.03
	0	1	1		0	1	0.1		0	1	0.09
	0	2	1		0	2	0.6		0	2	0.18
<i>step-1</i> →	1	0	1	<i>step-2</i> →	1	0	0.21	<i>step-3</i> →	1	0	0.02
	1	1	1		1	1	0.06		1	1	0.06
	1	2	1		1	2	0.12		1	2	0.12
	2	0	1		2	0	0.50		2	0	0.05
	2	1	1		2	1	0.15		2	1	0.15
	2	2	1		2	2	0.30		2	2	0.30

Now we can prove that the algorithm indeed converges and computes the unbiased reconstruction. Convergence is obvious because the algorithm is not iterative.

Let  $f = *^k_k f$  as computed by the basic join procedure [9] or the unified join procedure

as described previously. These procedures use complete information. The algorithm will terminate after one iteration. Then for any state  $\alpha$ ,

$$f_S(\alpha) = \prod_k f({}^k\alpha) \text{ where } {}^k\alpha \in [F \downarrow {}^kV], {}^k\alpha < \alpha, \quad (2.8)$$

and  ${}^k f({}^k\alpha)$  denotes the basic probabilities. There are no conditional probabilities when dealing with U-structures. Now, let  $f$  be the reconstruction computed by our U-algorithm. Then any  $f(\alpha)$  can be expressed as

$$f(\alpha) = \prod_k f({}^k\alpha) \prod_k f_c({}^k\alpha). \quad (2.9)$$

Note that substates involving  ${}^k f_c({}^k\alpha)$  are essentially all of those who were not part of the independent information. Thus  $f(\alpha)$  and  $f_S(\alpha)$  are the same. Because  $f_S$  is an unbiased reconstruction, so is  $f$ . This concludes the following theorem.

**Theorem 2.3** Given a probabilistic U-structure system  $S$ , U-algorithm determines the unbiased reconstruction from  $S$  using only the independent information.

#### 2.4.2 Concluding Remarks

We have provided a mechanism to employ independent information for U-structures and given a proof to this effect. Our U-algorithm is basically a partial join procedure which can greatly reduce the number of equations employed in the process of reconstruction when there are a large number of uncoupled subsystems. Thus, this result is important in reducing the storage requirements, which has been a major concern in Computer Science.

## CHAPTER 3

# RECONSTRUCTABILITY ANALYSIS OF POSSIBILISTIC SYSTEMS

### 3.1 Introduction

In the previous chapter, we discussed a few results for probabilistic reconstructability analysis, and provided some new results. A fair amount of work has been done on reconstruction of probabilistic behavior functions. In this chapter we shall aim at the reconstruction of possibilistic behavior functions. The principle of maximum uncertainty for possibilistic systems introduced by Higashi and Klir serves as the counterpart to the principle of maximum entropy for probabilistic systems. It satisfies not only the properties of symmetry, expansibility, subadditivity, additivity, normalization and continuity, as does its probabilistic equivalent, but also some additional properties, such as monotonicity [26, 45].

Thus, the principle of maximum uncertainty can be used to compute the unbiased reconstruction from possibilistic structure systems. Cavallo and Klir [11] consider the reconstruction of possibilistic behavior systems. Using the principle of maximum ambiguity, which was later revised to the principle of U-uncertainty, they introduce methods for computing the unbiased reconstruction and reconstruction families of possibilistic structure systems. They prove that a possibilistic join procedure, the one similar to the probabilistic one, computes the unbiased reconstruction and that there is no need to employ an iterative procedure for the structure systems with loops. Higashi, et. al. [27] demonstrate that the reconstruction family of a given structure system is equivalent to the set of solutions of a special kind of fuzzy relation equations. The partially ordered solution set contains the minimal solutions and the unique maximum solution. Identifying these maximum and minimal elements only suffices to determine

the whole reconstruction family. They also develop a measure of reconstruction uncertainty in order to define an identifiability quotient. This identifiability quotient describes the degree of confidence in identifying a single overall system for a given structure system.

In real life, there may be situations when information is not available in its entirety, or it may be cost prohibitive to observe all the states in an experiment. For example, in genetics, where there are dominant and recessive genes, some states are readily observed whereas the knowledge of others requires expensive testing. Therefore, the task of working with limited information is of paramount concern in these circumstances. It is the objective of this chapter to study such problems. In this chapter, we first present a method for determining the unbiased reconstruction for possibilistic behavior functions using partial information. Then we describe an algorithm to determine the reconstruction family of possibilistic systems. It is important to note that the problem of computing reconstruction family of possibilistic systems can be translated into the problem of solving a set of a special kind of fuzzy relation equations [27]. We use this equivalence in solving the identification problem.

### 3.2 Reconstruction of Possibilistic Systems

The discipline of reconstructability analysis is well developed for probabilistic behavior functions. Of greater significance has been Jones' work [33-41]. He has introduced the concept of the null extension, and thus provides a mechanism to enable reconstructability analysis to proceed without the availability of information in its entirety. We use the similar concept to study possibilistic systems with limited information.

Let  $B = (V, W, s, A, Q, f)$  be a possibilistic behavior function as defined in (1.1). Let  $S = \{ {}^k B \}$  be the reconstruction hypothesis as defined in (1.2). All the symbols have the same meaning as defined earlier except that the functions  $f$  and  $\{ {}^k f \}$

are possibilistic behavior functions. Similar to the probabilistic systems, the possibilistic systems also satisfy covering condition, irredundancy condition and local consistency condition as defined in (1.5), (1.6) and (1.7) respectively. Let  $Z \subset V$ , then  $[f \downarrow Z]$  is the projection of a possibilistic function in that it involves only those variables which are in set  $Z$ . Formally,  $[f \downarrow Z]$  is defined as

$$[f \downarrow Z]: \prod_{v_i \in Z} \mathcal{S}(v_i) \rightarrow [0, 1] \quad (3.1a)$$

such that

$$[f \downarrow Z] = \max_{\alpha \succ \beta} f(\alpha) . \quad (3.1b)$$

A justification of this definition has been provided in [69, 70, 71].

Evaluation of the reconstruction hypotheses for possibilistic systems relates to either of the following two problems, depending on whether or not the overall system is known. These two problems are the reconstruction problem and the identification problem. The identification problem requires the computation of the reconstruction family of  $B$  denoted by  $B_S$ . To choose a single system function  $f_S$  from  $B_S$  requires some additional assumptions, depending on whether some extra information about the overall system under investigation is available or not. In the event of not having such information,  $f_S$  should be maximally non-committal except for the following condition:

$$[f_S \downarrow {}^k V] = {}^k f \text{ for all } k \in N_q . \quad (3.2)$$

For a probability distribution, this amounts to saying that the set  $\{f_S(\alpha) \mid \alpha \in A\}$  must have maximum entropy subject to the above constraint. The principle of maximum entropy is well established and has been derived axiomatically as a general principle of inductive inference [29, 63]. The principle of maximum entropy determines a hypothetical probability distribution from the available partial information about a

probability distribution. This hypothetical distribution contains all the available information but no unsupported information and is unbiased and most likely to occur.

Higashi, et. al. [26] develop the possibilistic counterpart of the principle of maximum entropy to define a suitable measure of uncertainty. This measure, called U-uncertainty, is given as:

$$U(f) = \frac{1}{l_f} \sum_{k=1}^{q-1} (l_{k+1} - l_k) \log_2 |c(f, l_{k+1})| \quad (3.3a)$$

or

$$U(f) = \frac{1}{l_f} \int_0^{l_f} \log_2 |c(f, l)| dl \quad (3.3b)$$

where

$$f = (\phi_i \mid i \in N_{|X|}), \quad (3.3c)$$

$$l_f = \max_i \phi_i, \quad (3.3d)$$

$$L_f = \{ l \mid (\exists i \in N_{|X|}) (\phi_i = l) \text{ or } l = 0 \} = \{ l_1, l_2, \dots, l_q \}, \quad (3.3e)$$

and

$$c(f, l) = \{i \in N_{|X|} \mid \phi_i \geq l\}. \quad (3.3f)$$

$L_f$  is called a level set of  $f$ ,  $c$  is called the  $l$ -cut function and  $c(f, l)$  is called an  $l$ -cut of  $f$ .

As discussed previously, U-uncertainty is a measure of uncertainty for possibilistic systems which satisfies some additional properties beyond those satisfied by entropy and can be used to justify the selection of a particular function from the reconstruction family in the context of possibilistic systems.



### 3.2.1 Reconstruction Algorithm

We first define the concepts of *null extension*, *minimal substate* and *partial reconstruction hypothesis*.

**Definition 3.1** Let  $\beta$  be a substate. Then  $\alpha \in A$  is said to be a null extension of  $\beta$  if  $\alpha > \beta$  and every variable of  $\alpha$  which does not occur in  $\beta$  has a zero value.

**Definition 3.2** The two substates are said to be equivalent if and only if they have same null extension.

**Definition 3.3** There may be more than one substate with the same null extension. The one with the least function value is said to be the minimal substate.

**Definition 3.4** If the subsystems related to a reconstruction hypothesis are not complete then the reconstruction hypothesis is said to be a partial reconstruction hypothesis.

We reiterate that the concept of a null extension divides the whole substate space into disjoint equivalence classes. No substate can occur in two equivalence classes. It is important to note that, in the context of probabilistic reconstructability analysis, the concept of null extension was used to generate independent information. Two states were said to be equivalent if they were in the same equivalence class. It suffices to work with only independent information. We must emphasize that in the context of possibilistic reconstructability analysis, the concept of a null extension does not necessarily generate independent information. Rather, it provides a tool to carry out reconstructability analysis in the absence of complete information.

**Corollary 3.1 :** Let  ${}^1f$  and  ${}^2f$  be two possibilistic behavior functions such that  ${}^1f: X_1 \times X_2 \rightarrow [0, 1]$ , and  ${}^2f: X_2 \times X_3 \rightarrow [0, 1]$ . Then their join  ${}^1f * {}^2f$  is a function  ${}^1f * {}^2f: X_1 \times X_2 \times X_3 \rightarrow [0, 1]$  such that

$$[{}^1f * {}^2f](\alpha, \beta, \gamma) = \min[{}^1f(\alpha, \beta), {}^2f(\beta, \gamma)] . \quad (3.4)$$

**Corollary 3.2:** Let  $f_S = *^k_k f$ . Then  $f_S$  is unbiased reconstruction implied by  $S$ .

Corollary 1 & 2 are due to [11]. Following is the possibilistic version of the reconstruction algorithm given by Jones [35].

Given a consistent reconstruction hypothesis in the form of equations  $\max_{\alpha > \beta} f(\alpha) = {}^k f(\beta)$  for all  ${}^k f(\beta)$  available. We obtain an unbiased reconstruction  $f_{S_{Partial}}$  using the following steps:

[1]  $f(\alpha) := 1$  for all  $\alpha$  in the system;

[2] for all  ${}^k f(\beta)$  available do the following computation:

$$\begin{aligned} f(\alpha) &:= f(\alpha) * {}^k f(\beta) \\ &= \min [f(\alpha \mid \beta), {}^k f(\beta)] \\ &= \min [f(\alpha), {}^k f(\beta)]; \end{aligned}$$

[3]  $f_{S_{Partial}} := f$ ;  $f_{S_{Partial}}$  is an unbiased reconstruction for the available information;

[4] stop.

We illustrate above algorithm using the following example. Given the following reconstruction hypothesis  $\{\{v_1, v_2\}, \{v_2, v_3\}, \{v_1, v_3\}\}$ ,

$\underline{v_1}$	$\underline{v_2}$	$\underline{{}^{12}f}$	$\underline{v_2}$	$\underline{v_3}$	$\underline{{}^{23}f}$	$\underline{v_1}$	$\underline{v_3}$	$\underline{{}^{13}f}$
0	0	0.8	0	0	0.8	0	0	0.8
0	1	0.5	0	1	0.7	0	1	0.7
1	0	0.0	1	0	0.8	1	0	0.8
1	1	0.8	1	1	0.5	1	1	0.5

following are the equivalence classes.

<u>Equivalence Class</u>	<u>Null Extension</u>	<u>Equivalence Class</u>	<u>Null Extension</u>
$\{ {}^1(0), {}^2(0), {}^3(0), {}^{12}(00), {}^{23}(00), {}^{13}(00) \}$	000	$\{ \}$	111
$\{ {}^3(1), {}^{23}(01), {}^{13}(01) \}$	001	$\{ {}^{13}(11) \}$	101
$\{ {}^2(1), {}^{12}(01), {}^{23}(10) \}$	010	$\{ {}^{12}(11) \}$	110
$\{ {}^1(1), {}^{12}(10), {}^{13}(10) \}$	100	$\{ {}^{23}(11) \}$	011

Choosing the minimal element from each equivalence class (if there are more than one minimal element then choosing any one of them), we get

$$S_{\text{Partial}} = \{ {}^{12}(00), {}^{13}(01), {}^{12}(01), {}^{23}(11), {}^{12}(10), {}^{13}(11), {}^{12}(11) \}.$$

The unbiased reconstruction determined by the above algorithm is

$$(f(000) = 0.8, f(001) = 0.7, f(010) = 0.5, f(011) = 0.5,$$

$$f(100) = 0.0, f(101) = 0.0, f(110) = 0.8, f(111) = 0.5).$$

We state, in view of Corollary 3.1 and Corollary 3.2, that the algorithm considers only the given information and no additional information. Thus the algorithm computes the unbiased reconstruction for the information employed. Then  $f_{S_{\text{Partial}}}$  is an unbiased reconstruction implied solely by  $S_{\text{Partial}}$ . However, we cannot say if  $f_S$  is the unbiased reconstruction for the whole system.

In order to prove that  $f_{S_{\text{Partial}}}$  is the unbiased reconstruction solely implied by  $S_{\text{Partial}}$  we give the following definition.

**Definition 3.5:** Let  $F = \{ f \mid f: A \rightarrow [0, 1] \}$ . Let  $\leq$  denote a partial ordering in  $F$  such that for each pair  $f_1, f_2 \in F$ ,  $f_1 \leq f_2$ , if and only if  $f_1(\alpha) \leq f_2(\alpha)$  for all  $\alpha \in A$ .

Given a partial reconstruction hypothesis  $S_{\text{Partial}}$  we assert that the reconstruction family of  $S_{\text{Partial}}$ ,  $F_{S_{\text{Partial}}}$  has a unique maximum  $f_{S_{\text{Partial}}}$  with respect to partial ordering  $\leq$  which can be determined by  $f_{S_{\text{Partial}}} = \bigstar_k {}^k f(\beta)$  for  ${}^k f(\beta)$  in  $S_{\text{Partial}}$ . Similar to

Higashi, et. al. [26], let  $f \in F_{S_{Partial}}$ . Then  $f(\alpha) \leq \min_j {}^j f(\alpha \downarrow {}^j V)$  for all  $\alpha$  and for all  ${}^j f(\alpha \downarrow {}^j V)$  in  $S_{Partial}$ . On the other hand, let there be some  $\alpha_0 \in A$  such that  $f(\alpha_0) > \min_j {}^j f(\alpha_0 \downarrow {}^j V)$ . Then there exists some  ${}^{j_0} f$  in  $S_{Partial}$  such that  $f(\alpha_0) > {}^{j_0} f(\alpha_0 \downarrow {}^{j_0} V)$ , which contradicts the very definition of projection function in possibilistic systems. Thus  $f \notin F_{S_{Partial}}$ . Thus  $f \in F_{S_{Partial}} \Rightarrow f \leq f_{S_{Partial}}$ . Now,  $f \leq f_{S_{Partial}} \Rightarrow \max_{\alpha > {}^j \alpha} f(\alpha) \leq \max_{\alpha > {}^j \alpha} f_{S_{Partial}}(\alpha)$  for all  ${}^j \alpha$  in  $S_{Partial}$ , or

$${}^j f({}^j \alpha) \leq [f_S \downarrow {}^j V]({}^j \alpha) \quad (3.5)$$

for all  ${}^j \alpha \in {}^j A$  in  $S_{Partial}$ .

For some  $\alpha \in A$  such that  $\alpha > {}^j \alpha$ ,  $f_S(\alpha) = \min_i {}^i f({}^i \alpha) \leq {}^j f({}^j \alpha)$  for all  ${}^i \alpha \in {}^i A$  in  $S_{Partial}$ . So  $f_S(\alpha) \leq {}^j f({}^j \alpha)$ . But  $[f_S \downarrow {}^j V]({}^j \alpha) = \max_{\alpha > {}^j \alpha} f_S(\alpha) \leq f_S(\alpha) \leq {}^j f({}^j \alpha)$ .

Therefore

$$[f_{S_{Partial}} \downarrow {}^j V]({}^j \alpha) \leq {}^j f({}^j \alpha). \quad (3.6)$$

From (3.5) and (3.6),  $[f_{S_{Partial}} \downarrow {}^j V]({}^j \alpha) = {}^j f({}^j \alpha)$  establishing  $f_{S_{Partial}} \in F_{S_{Partial}}$ . This concludes the proof of the following theorem.

**Theorem 3.1:** If  $F_{S_{Partial}}$  is non-empty, then  $F_{S_{Partial}}$  has a unique maximum  $f_{S_{Partial}}$  with respect to partial ordering  $\leq$  which can be determined by using partial join ( as described by the algorithm ).

### 3.3 Reconstruction Families of Possibilistic Systems

A reconstruction family of a given structure system can be considered equivalent to the set of solutions of a special type of fuzzy relation equations. The solution set thus obtained is partially ordered and contains both minimal solutions and unique maximum solutions. It suffices only to identify the maximum and minimal elements in

order to determine the whole reconstruction family. This was the idea used by Higashi, et. al. [27] in identifying the reconstruction family of possibilistic systems. In this section, we extend the research on the same line by providing a method for partial reconstruction hypotheses. We shall discuss how to determine the reconstruction families of the partial reconstruction hypothesis using this approach.

A possibilistic measure is a special kind of fuzzy measure which is applicable only to finite sets and some special types of infinite sets [27,56,65]. However, we are concerned here with finite sets only. Let  $S_{\text{partial}}$  be a partial reconstruction hypothesis. Then all  $f \in F_{S_{\text{partial}}}$  can be determined by solving the set of simultaneous equations

$$\max_{\alpha > {}^j\alpha} f(\alpha) = {}^j f({}^j\alpha) \quad (3.7a)$$

for all  $\alpha$  in the system and for all  ${}^j\alpha$  in  $S_{\text{partial}}$ , along with the constraint

$$0 \leq f(\alpha) \leq 1. \quad (3.7b)$$

Equation (3.7a) can be expressed as

$$\max_{\alpha \in A} \min(f(\alpha), \delta_{\alpha, {}^j\alpha}) = {}^j f({}^j\alpha) \quad (3.8a)$$

for all  $\alpha \in A$  and  ${}^j\alpha$  in  $S_{\text{partial}}$ , where

$$\delta_{\alpha, {}^j\alpha} = \begin{cases} 1 & \text{if } \alpha > {}^j\alpha \\ 0 & \text{otherwise.} \end{cases} \quad (3.8b)$$

This problem can now be translated in terms of fuzzy relation equations as follows. Let  $p, q, r$  be fuzzy binary relations defined as [16,17,27,61]  $p: X \times Y \rightarrow [0, 1]$ ,  $q: Y \times Z \rightarrow [0, 1]$ ,  $r: Z \times X \rightarrow [0, 1]$ , and let “ $\circ$ ” be max-min composition. Then, the general form of a fuzzy relation equation can be written as

$$poq = r. \quad (3.9)$$

Thus, for all  $x \in X$  and  $z \in Z$ ,

$$p \circ q(x, z) = \sup_{y \in Y} \min(p(x, y), q(y, z)) . \quad (3.10)$$

Because  $X, Y, Z$  are finite sets, functions  $p, q, r$  can be mapped onto their respective matrices as follows:  $p = (p_{ij}), q = (q_{jk}), r = (r_{ik})$ , where

$p_{ij} = p(x_i, y_j), q_{jk} = q(y_j, z_k), r_{ik} = r(x_i, z_k)$  and  $x_i \in X, y_j \in Y, z_k \in Z$ . Now for each pair  $(i, k)$  we can write

$$r_{ik} = \max_j \min(p_{ij}, q_{jk}) . \quad (3.11)$$

Now we define

$$p = (p_1, p_2, \dots, p_{|A|}), p_i = f(\alpha_i), i \in N_{|A|} , \quad (3.12a)$$

$$Q = \begin{bmatrix} q_{11} & q_{12} & \cdot & \cdot & \cdot & q_{1|P_\phi|} \\ q_{21} & q_{22} & \cdot & \cdot & \cdot & q_{2|P_\phi|} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ q_{|A|1} & q_{|A|2} & \cdot & \cdot & \cdot & q_{|A||P_\phi|} \end{bmatrix} \quad (3.12b)$$

where

$$P_\phi = \{ \phi f(\beta_1), \phi f(\beta_2), \dots, \phi f(\beta_{|P_\phi|}) \} , \quad (3.12c)$$

$$q_{ik} = \begin{cases} 1 & \text{if } \alpha_i > \phi \beta_k \\ 0 & \text{otherwise} \end{cases} \quad (3.12d)$$

and

$$r = (r_1, r_2, \dots, r_{|P_\phi|}) . \quad (3.12e)$$

Note that  $\phi f$  are subsystem functions in  $P_\phi$ . Now equation (3.8a) can be rewritten as

$$\max_{i \in N_{|A|}} \min(p_i, q_{ik}) = r_k \quad (3.13a)$$

for all  $k \in N_{|P_e|}$  which essentially is

$$p_o Q = r. \quad (3.13b)$$

Now for the given example, we have the following representation of  $p_o Q = r$ ,

$$\begin{bmatrix} p_0, & p_4, & p_2, & p_1, & p_6, & p_3, & p_5, & p_7 \end{bmatrix} o \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} = [0.8, \quad 0.0, \quad 0.5, \quad 0.7, \quad 0.8, \quad 0.5, \quad 0.5].$$

The system of equations given above can further be simplified by eliminating those columns from  $Q$  which correspond to 0 values in  $r$ -vector. The solution to the above set of equations is  $p_0 = 0.8, p_1 = 0.7, p_4 = 0.0, p_5 = 0.0, p_6 = 0.8, p_7 = 0.5$ , and either  $p_2 = 0.5$  and  $p_3 \leq 0.5$  or  $p_3 = 0.5$  and  $p_2 \leq 0.5$ . Obviously, the reconstruction family has infinite number of elements. Only one of them is maximum represented by  $p_2 = 0.5$  and  $p_3 = 0.5$ . Two of them are minimal elements represented by  $p_2 = 0.0$  and  $p_3 = 0.5$  and represented by  $p_2 = 0.5$  and  $p_3 = 0.0$ .

### 3.3.1 Concluding Remarks

We have expressed the problem of determining the reconstruction family from a partial reconstruction hypothesis in terms of max-min fuzzy relation equations. Matrix  $Q$  is in the form of a lower triangular matrix which can greatly simplify the computation. By eliminating the columns corresponding to zero values in the  $r$ -vector, we can further simplify the computation. Details and proofs of this max-min approach are listed in Higashi, et. al. [27].

## CHAPTER 4

### **RECONSTRUCTABILITY APPROACH TO AUTOMATED RULE LEARNING**

Reconstructability analysis resembles the process of learning in the sense that it discovers general patterns in the data during the process of system reconstruction. Based on this fact, our approach to generalized rule induction from databases is unique in that it should work for most data covered by the framework of RA and k-systems. It is especially appropriate for expert systems domains where the data is intrinsically nominal, because RA assumes no structure in the data and regards them as categories. The decisions made by experts exhibit a great amount of subjectivity, uncertainty and arbitrariness. This is why there are times when the experts find themselves incapable of justifying their own correct decisions. It is our understanding that RA can be of tremendous utility in analyzing problems in this type of domain.

Our objective is to introduce a new concept for automated rule learning using the state/substate paradigm of RA. Because expert systems deal mostly with nominal data, the RA methodology seems to be very promising for this type of application. We refer to it as the reconstructability approach to rule learning. Since the problem of learning rules from databases is viewed as a reconstruction problem, we also call it learning by reconstruction.

Unlike classical or classification based techniques, the reconstructability approach does not induce any model on the data or on the nature of the solution, nor does it make any extraneous or erroneous assumptions regarding the data. The potential advantages attributed to this approach include the ability to select rules of any order and to learn multiple concepts, the explicit specification of unknown attribute values, the ability to deal with the noisy data effectively, the ability to work with the



partial information, and the avoidance of any strong biases in the process of generalized rule induction.

#### 4.1 Introduction

The induction, or the learning from examples is simply the search for hypotheses in some hypotheses space to account for a set of instances in some instance space. For the purpose of the learning approach introduced in this paper, the hypotheses space is restricted to the space of conjunctions of attribute values. Generally speaking, automated rule induction algorithms can be characterized by two learning paradigms, namely, symbolic manipulation techniques and statistical inference techniques [15]. Symbolic learning algorithms use a bottom-up approach that involves incremental generalization of the specialized hypotheses. Our algorithm and most statistical algorithms use a top-down approach that involves the specialization and the refinement of more general hypotheses in order to improve the goodness-of-fit. Unlike bottom-up approaches, the top-down approaches are generally characterized by the faster convergence rates and a higher degree of robustness [64].

Consider the problem of inferring the general rules (patterns) from the given databases. The knowledge acquisition bottleneck in obtaining rules directly from the experts is the central motivation to consider this type of problem. Therefore, for domains where a sample database already exists, the task of automated rule learning is of paramount concern. For instance, consider a company or an organization that wishes to develop a rule-based expert system for the domain to which the data is applicable. This expert system can be used for, for example, for prediction, diagnosis or simulation. How to use this existing database to effectively and automatically derive rules for this expert system is the topic of our concern here. The KRULE algorithm introduced in this chapter should lead us to learn the rules for one or more concepts.

## 4.2 Other Approaches to Learning

In this section, based on [12, 14-15, 18-19, 52-53, 57-60, 64], we outline the different approaches to learning and discuss the motivation behind pursuing our approach. As mentioned earlier, automatic rule induction algorithms can be characterized by two learning paradigms, namely, symbolic manipulation techniques and statistical inference techniques [15, 64]. Mitchell's version space approach [53] and Michalski's AQ [52] algorithms are the well known examples of symbolic learning, and are characterized by the implicit assumption of a Baye's misclassification rate of zero. These algorithms examine the examples sequentially, that is, one by one, and refine the rule space until a set of general classification rules covering the seen examples is achieved. These approaches are not computationally feasible and fall short of handling noise in the data. The first drawback is due to the exponential time complexity of the algorithms, and the second drawback lies in the implicit assumption that each attribute can be perfectly classified in terms of the other attributes. Most symbolic learning algorithms suffer from this implicit assumption of a zero value of Baye's misclassification rate [21, 25, 67]. On the other hand, the statistical techniques exploit average properties of the data set. Most learning algorithms using statistical inference lack the flexibility because either they assume a parametric statistical model on the data (such as multivariate analysis, factor analysis, or discriminant analysis) or they force a structure on the nature of solution (such as tree induction algorithms).

One of the widely used learning algorithms is Quinlan's ID3 algorithm [57, 59]. ID3 derives classification decision trees from the sample data. Despite their widespread popularity, decision trees are sequential decision algorithms which do not go well with the data driven approach of expert systems [64]. Rule based systems are data driven, that is, any set of input data can possibly be used to begin the process. However, trees always begin with the attribute associated with the root node. Rule based systems can handle the missing attribute information, while trees are not designed for

this purpose. Further, rule based systems enjoy the benefit of explicit knowledge representation inherent to the production rules, while trees are hard to understand for the users. Conclusively, rules provide greater flexibility in representation than trees, especially in the context of expert systems.

The problem of rule induction from databases is referred to as generalized rule induction [64]. A few attempts have been made to solve these problems. Gaines and Shaw [19] have proposed an algorithm, called ENTAIL, which is designed to handle subjective data and is based on the fuzzy logic. ENTAIL derives rules from a repertory grid and outputs the set of most informative rules. The main idea behind ENTAIL is to transform the subjective repertory grid numbers and then use them to obtain a measure of the information content of the associative rules. Ganascia's approach [20] for rule induction is a heuristic one, rather than algorithmic, that does not employ any specific measure of rule goodness. Cheeseman [13] uses a Bayesian approach to rule induction based on the probabilistic measures.

Quinlan [58] introduces a classification based scheme of transforming ID3 induced trees into the production rules. This scheme has two major drawbacks. First, this scheme is classification based, that is, based on the assumption of a zero value for Baye's misclassification rate. Second, the general optimality of tree transformation techniques for performing rule induction is doubtful. Similar to Quinlan's approach, Cendrowska's PRISM algorithm [12] is also classification based and uses information theoretic measures. This scheme also suffers from the disadvantages inherent in the classification based schemes.

Recently, Clark and Niblett [14] describe a rule induction algorithm, called CN2. This algorithm is similar to PRISM in that it directly employs a measure of rule goodness in order to search for classification rules. CN2 has two drawbacks. First, it allows for internal disjunctives and thus occupies a larger hypothesis space. Second, it

produces a set of rules in the form of a decision list [60]. Because a decision list is a form of decision tree, CN2 suffers from the same limitations as trees.

Following are the shortcomings of CN2 and PRISM. Both of them do not include a priori probability term in their rule measures [12, 14, 64]. As stated in [64], inclusion of a priori probability is a necessary requirement in a scheme which performs generalized rule induction, for this allows comparisons of the competing hypotheses for the same concept as well as hypotheses for different concepts. From an information theoretic point of view, the rarer an occurrence of an event, the more valuable is the information confirming its occurrence. The ability to rank competing hypotheses for multiple concepts is essential for a learning algorithm in a resource constraint environment. This is the central theme of a recent paper by Smyth and Goodman [64]. They use a probabilistic approach to generalized rule induction from databases using information theoretic measures and a priori probability term. This algorithm is basically aimed at solving the problems in the probabilistic domain and therefore is not general enough to cover a variety of applications.

Our approach to generalized rule induction is more general because it should work for most data using the tools and techniques of k-systems. The state/substate paradigm of RA has led to a departure from classical statistical approaches, and thus, has provided new insights into the structure and dynamics of the systems. Contrary to RA, classical statistical approaches emphasize variables [40]. Concepts of state and substate in RA have no equivalents in classical framework. A state viewed apart from its occurrence is important everywhere in RA, illustratively, in null extensions or in evaluating the cognitive contents of system substates [37]. RA is as good as the data it is given. It does not force any structure on the data. On the contrary, statistical techniques induce some kind of model on the data and hence introduce extraneous information [40].

### 4.3 Measures of Cognitive Contents of a Rule

In this section we discuss different measures of the cognitive contents of a rule and introduce a new measure. We refer to this new measure as a  $K$ -measure. This measure quantifies the cognitive contents of a rule relative to the system it represents. Throughout this text, we use the example of a rule and the related concepts given by Smyth and Goodman [64]. Consider the following example of a rule.

$$\text{If } Y = y \text{ then } X = x \text{ with probability } p. \quad (4.1)$$

Here  $X$  and  $Y$  are two attributes or two dimensions in the state space and  $x$  and  $y$  are their values. Let  $X$  and  $Y$  be discrete random variables. Let  $X = x$  represent a single value assignment statement. Let  $Y = y$  represent a single expression or conjunction of expressions. A rule corresponds to particular condition or input event  $Y = y$ .

Let  $f(X; Y = y)$  be the instantaneous information that event  $Y = y$  provides regarding  $X$ . In other words, this is the information received about  $X$  knowing that  $Y = y$  is true or has occurred. Shannon [62] states that the average information from all rules should be consistent with the standard definition for average mutual information. This amounts to the following equality

$$E_y[f(X; Y = y)] = I(X; Y). \quad (4.2)$$

Here  $E_y$  denotes the expectation with respect to the random variable  $Y$ . Blachman [4] proposes the following two measures, called the  $i$ -measure and  $j$ -measure, in order to define  $f(X; Y = y)$ . The  $i$ -measure is given as

$$i(X; Y = y) = \sum_x p(x) \cdot \log \frac{1}{p(x)} - \sum_x p(x | y) \cdot \log \frac{1}{p(x | y)}. \quad (4.3)$$

As is evident from the above formula, the  $i$ -measure represents the difference in the *a priori* and *a posteriori* entropies of  $X$ . As discussed by Smyth and Goodman, there are problems with the  $i$ -measure which make it inappropriate as a basic measure

of cognitive content of a rule [23, 64]. For instance, the  $i$ -measure may take negative values. Besides,  $i(X; Y = y)$  can become zero even if  $p(x | y) \neq p(x)$ . For example, if  $p(x | y) = p(\bar{x})$  and  $X$  is a binary variable, then  $i(X; Y = y) = 0$ . This amounts to stating that there is no change in entropy, though we have received information about  $X$ . This phenomenon is referred to as an *information paradox* in the literature. Furthermore, the  $i$ -measure is incapable of distinguishing between individual events. For example, let  $X$  be an  $n$ -valued variable where  $X = x_1$  is likely to occur a priori with probability  $p(x_1) = 1 - \varepsilon$ . Assume that other values of  $X$  are equally unlikely to occur with probability  $\varepsilon/(n - 1)$ . In this situation, a conditional probability  $p(x | y)$  represents a rule that predicts the relatively rare event  $X = x_r$  for some  $r$ , and is thus significant to the analysis. However, the  $i$ -measure fails to capture this information and would yield zero information for such an event.

Considering the drawbacks of the  $i$ -measure, Smyth and Goodman [64] propose the  $j$ -measure as the information content of a rule. The  $j$ -measure is given as

$$j(X; Y = y) = \sum_x p(x | y) \cdot \log \frac{p(x | y)}{p(x)}, \quad (4.4a)$$

$$\text{or } j(X; Y = y) = p(x | y) \cdot \log \frac{p(x | y)}{p(x)} + (1 - p(x | y)) \cdot \log \frac{1 - p(x | y)}{1 - p(x)}. \quad (4.4b)$$

The  $j$ -measure represents the average mutual information between the events  $x_i$  and  $y$ , expectation being taken with respect to the posteriori probability distribution of  $X$ . The  $j$ -measure has several advantages as compared to the  $i$ -measure. Blachman [4] proves that the  $j$ -measure is a unique non-negative information measure. Smyth and Goodman [64] have found that the  $j$ -measure satisfies a variety of desirable properties, including appropriate limiting properties. For instance, as the transitional probability (that is, conditional probability) approaches one, the information content of the rule approaches  $\log(1/p(x))$  which is the self information of the event at the right hand side

of the rule. The  $j$ -measure can be regarded as a special case of Shore and Johnson's cross entropy [63]. It can also be viewed as discrimination, as described by Kullback [49] and Blahut [5]. Discrimination is a measure of information theoretic similarity between two probability distribution. In this respect the  $j$ -measure describes the dissimilarity between a priori and a posteriori beliefs about  $X$ . The higher the degree of dissimilarity, more useful are the rules.

We now define the  $J$ -measure as follows,

$$J(X; Y = y) = p(y) \cdot j(X; Y = y). \quad (4.5)$$

As discussed by Smyth and Goodman, the  $J$ -measure is an average measure of cognitive contents of a rule as it does not take into consideration the instantaneous information from the other  $Y$  terms. Whereas the average measure pertains to the average value of the rule information content, the instantaneous measure can be used to rank rules after the event  $Y = y$  has occurred. The former can be used for learning while the latter can be used for forward chaining in drawing inferences in rule based systems [64]. Note that the  $J$ -measure implicitly ignores the instantaneous information from the other  $Y$  terms. Though this may be consistent, in the context of learning in a resource constrained environment based on the assertion that each rule must be significant in its own right, it is of utmost significance to discover how a rule interacts with the other rules and what is its contribution to the overall knowledge of the system it represents. This has been the motivation for introducing our  $K$ -measure. Succinctly, there are three issues in choosing rules,

- instantaneous information contents of a rule, represented by the  $j$ -measure,
- average information contents of a rule, represented by the  $J$ -measure, and
- amount of information a rule contributes to the system it represents.

A rule may contain enough information in absolute sense but may not have any

importance for the system as a whole. Therefore, the cognitive contents of a rule need to be explored in relative terms. In order to cope with this situation, we introduce a new measure of information based on the principles of reconstructability theory. Jones [37,41] has described a measure which estimates the cognitive contents of a system substate. Let this measure be referred to as the  $k$ -measure, defined as

$$k(X; Y = y) = k(\beta) = {}^mk(\beta) \cdot \log \frac{{}^mk(\beta)}{{}^mk'(\beta)} + (1 - {}^mk(\beta)) \cdot \log \frac{1 - {}^mk(\beta)}{1 - {}^mk'(\beta)} \quad (4.6)$$

where substate  $\beta$  can be expressed as  $(Y = y) \wedge (X = x)$ ,  ${}^mk$  are true subsystem functions (that is, probabilities) and  ${}^mk'$  are approximate subsystem functions,  $m$  being the number of total subsystems.

We reiterate that a state or substate is simply a conjunction of attribute values or variable values. The functions  $k$ ,  $k'$ ,  ${}^mk$  and  ${}^mk'$  represent the probabilities (normalized function values, in the general case [41]) of states and substates, respectively. Note that the  $k$ -measure will choose the substate  $\beta$  that will add the most information to the system being reconstructed. Thus, the  $k$ -measure is essentially a measure of instantaneous information of the substate (rule) relative to the system. Now we define our  $K$ -measure, the average measure of cognitive contents of a rule with respect to the system. The  $K$ -measure is given as

$$K(X; Y = y) = p(y) \cdot k(X; Y = y). \quad (4.7)$$

Illustratively, if  $(v_1 = 0, v_2 = 1, v_3 = 0, p = 0.09)$  is an item in the database (or system) then  $p(y)$  is simply the probability of substate  $(v_1 = 0) \wedge (v_2 = 1)$ , and  $k(X; Y = y)$  is the measure of instantaneous information for the state  $(v_1 = 0) \wedge (v_2 = 1) \wedge (v_3 = 0)$ .

#### 4.4 KRULE: Rule Induction Algorithm

The problem of generalized rule induction is the problem of finding a hypothesis to fit some given data. Simplicity of the hypothesis and the goodness-of-fit between



the hypothesis and the data are two primary criteria in the evaluation of the hypothesis [1,64]. Similar to the  $J$ -measure, the  $K$ -measure is the product of terms  $p(Y = y)$  and  $k(X; Y = y)$ . The first term  $p(Y = y)$  is the probability of occurrence of the hypothesis while the second term  $k(X; Y = y)$  represents the contribution of state  $x \wedge y$  to the data. Therefore, the former can be regarded as a measure of simplicity, the latter can be regarded as a measure of goodness-of-fit. In our case, we refer to the latter as a measure of goodness-of-representation. Thus, our  $K$ -measure should possess a combined meaning for the simplicity and goodness-of-representation of a given rule.

Following is an outline of the rule induction algorithm. We refer to it as the KRULE algorithm. In order to be able to learn a concept  $X$ , we follow the following steps:

- [1] search for the states which include  $X$ ;
- [2] compute  $p(Y = y)$  and compute  $k(X; Y = y)$  using (4.6);
- [3] compute  $K(X; Y = y)$  using (4.7);
- [4] among the possible rules choose one which has the highest value of  $K$ -measure;
- [5] if more rules needed, go to step [2];
- [6] stop.

In the above algorithm, observe that it is possible to learn multiple concepts. Once  $k(X; Y = y)$  is computed,  $K(X; Y = y)$  can be simultaneously computed for other concepts. Note that this approach requires the computation of a priori probabilities. Where a large amount of sample data is available, the computation can be performed using various available techniques [22]. Another limitation of this approach is that it is primarily aimed to work at discrete data. Special treatment of data such as clustering is required under these circumstances. Technique are available for the transformation of data [68].

#### 4.5 Advantages of the KRULE Approach

Besides being suitable to expert system domains, there are certain undisputed advantages attributed to this approach. This approach allows us to choose rules of any order. For example, we can specify that we are looking for third order rules only, or that we are looking for rules of order three or less. To achieve this, we need to choose only states carrying three or less variables during the system reconstruction. This offers a high degree of freedom and flexibility in determining the desired level of generalization (or specialization) and in shaping the knowledge base suitable to one's requirements. Of course, as is always the case, there is a trade-off between accuracy and the level of generalization.

Another specific advantage of this scheme is the explicit specification of unknown attribute values. As the data dealt with is categorical, if some attribute value is not specified in a tuple (missing attribute value), we can arbitrarily assign a new value to this attribute, say  $u$  for unknown. Later on, an occurrence of this attribute value in any rule will explain the importance of the knowledge of  $u$ . A significant advantage of this approach is its ability to deal with noisy data effectively. This is because RA searches for general patterns and disregards redundant information. Also, it is fair to state that an unbiased reconstruction solution provided by RA methodology will avoid any strong biases in the process of generalized rule induction.

Finally, it is possible for our learning algorithm to work for insufficient and incomplete data. The promise to do so is shown in RA and k-systems methodology [33-45]. Given partial data, we can use k-systems theory and reconstructability theory, under most circumstances, to reconstruct the original system (unbiased reconstruction), and then to employ stated learning algorithm.

## 4.6 Concluding Remarks

In this chapter, we have presented KRULE, a rule induction algorithm based on a reconstructability theoretic measure called the  $K$ -measure. The  $K$ -measure determines the information content of a rule relative to the given data. Based on reconstructability theory, the  $K$ -measure is the *unbiased* measure of the cognitive contents of a rule, which does not include any unsupported information. This approach is not only good for probabilistic data but should encompass, in most cases, a wide range of problems using reconstructability analysis for general systems.

Finally, a word on the complexity of our algorithm. Search is inherently exponential. Since the left side of the rule allows conjunctions only, the search space is drastically reduced. Furthermore, bounds expressed by Smyth and Goodman [64] on the  $J$ -measure can be used to reduce the search and proceed with specialization and generalization as needed. However, we shall not elaborate on these issues here. Like the  $J$ -measure, the  $K$ -measure is a non-negative measure of the cognitive content of a rule relative to the system which satisfies Shannon's requirement (4.2). However it should perform better than the  $J$ -measure in terms of knowledge representation, especially when there is a higher degree of interaction between various data components. Furthermore, as noted in early chapters, the concept of a null extension can be used to divide the state space into equivalence classes, and the process of reconstruction can be performed concurrently in order to ease the time factor.

## CHAPTER 5

### CONCLUDING REMARKS

#### 5.1 Summary and Conclusions

The two problems in reconstructability theory are referred to as the reconstructability problem and the identification problem. The reconstructability problem relates to the process of reconstructing a given system under a given criterion from the knowledge of its subsystems and, during this process, identifying the subsystems that are important in the reconstruction. The identification problem permits the identification of an unknown system from the knowledge of its subsystems. The solution procedures associated with these two problems are referred to as Reconstructability Analysis, abbreviated as RA. Advent of RA has intensified the research efforts on system studies. The point is to view the systems and subsystems as entities interrelated by some mathematical criterion. The objective of this research has been to study the process of system reconstruction and to apply the results from reconstructability analysis to the problem of automated rule learning. We have furthered the realm of RA by studying it in the context of probabilistic systems, selection systems and possibilistic systems. Based on the RA methodology and the k-system framework, we have introduced a new concept to automated knowledge acquisition from databases.

In the first chapter we have introduced the preliminary concepts in the reconstructability theory and machine learning such as systems, subsystems, states, sub-states, reconstructibility problem, identification problem, k-systems, automated knowledge acquisition, expert systems, etc.. We have elaborated on the evaluation of reconstruction hypotheses, have described the significance of unbiased reconstruction, and have emphasized the usefulness of k-systems and RA methodology.

In the second chapter we have extended the previous work on reconstructability analysis for probabilistic systems and selection systems to generate better algorithms for determining the unbiased reconstruction and reconstruction families. We have provided a generalization to Jones' method [33] in the sense that the computation of reconstruction families is not limited by the choice of substates to a particular ordered set  $P_\phi$ . Any substate belonging to the same equivalence class as the member of  $P_\phi$  can be employed in the process of reconstruction. Reconstruction families are determined in the form of matrix equations  $MQ = W$  representing set of linearly independent algebraic equations. The coefficient matrix  $M$  is an upper triangular matrix which greatly reduces the computation.

Further, we have given a method to compute the unbiased reconstruction for a selection system based on possibilistic measures. Furthermore, for a probabilistic system, we have proposed an algorithm to compute the unbiased reconstruction. This algorithm makes computation simple by avoiding the explicit computation of probabilities. Moreover, we have given an algorithm to compute the unbiased reconstruction for a special class of systems, called a U-structure. A proof for computing unbiased reconstruction for U-structures is given when only independent information is employed. It is really appealing to be able to work with partial information, because at times it is cost prohibitive to observe all states.

In the third chapter we have provided algorithms for computing unbiased reconstruction and reconstruction families of possibilistic systems using partial information. To this end, we have introduced the concept of the partial reconstruction hypothesis. A possibilistic version of the probabilistic algorithm is provided to compute the unbiased reconstruction, and the reconstruction families have been determined by transforming the possibilistic system constraints into max-min fuzzy relation equations. The coefficient matrix is lower triangular which greatly simplifies the computation.

Finally, based on the tools and techniques of the k-systems framework, we have introduced a new measure of the cognitive contents of a rule. We refer to this measure as a K-measure. When a large sample database exists, it is highly desirable to employ automated knowledge acquisition to learn important concepts. Based on reconstructability theory, our approach to rule learning from databases is unique in that it should work for most data covered by the framework of RA. In particular, it is very appropriate for expert systems domains where the data is inherently nominal. Unlike classical or classification based techniques, our approach does not induce any model on the data or on the nature of the solution, nor does it make any extraneous or erroneous assumptions regarding the data. Our approach to generalized rule induction from databases should reveal the broader spectrum in the data by concentrating information into a few descriptive rules.

## 5.2 Problems and Potentials

In this dissertation, we have extended the work on possibilistic systems using partial information and have determined the reconstruction families in terms of fuzzy relation equations. It is desirable to extend the reconstructability analysis for fuzzy measures such as plausibility, necessity or credibility measures [45] or for a more general class of fuzzy functions.

We conclude by emphasising that the RA methodology has great potential to support applications in a variety of disciplines [42]. We should extend the RA methodology to suit the requirements of a particular learning process. For example, in the context of expert systems, we may have a function whose values are drawn from a nominal data set. We will refer to this function as a nominal function and the system as a nominal system.

Consider a situation in a medical domain where the system function may represent a disease that is based on given attributes. The disease may be sinus headache,

cholera, or flu. It is not always possible to quantify such function values, or to impose an ordering upon them, which is necessary to investigate the system as a probabilistic, possibilistic, or non-linear system. Currently, no RA methodology is available to handle this kind of problem effectively.

The Following is an example of a nominal system:

$v_1$	$v_2$	$f$
0	0	flu
0	1	cholera
1	0	sinus headache
1	1	cholera

One way to approach this problem is to regard a nominal function as a variable, thereby introducing one more variable to the set of variables, and simply observe the occurrence or non-occurrence of the aggregate states. Theoretically, this transforms the original system into a selection system. Appropriate solution procedures can now be developed based on the methodology of RA and k-systems.

A selection system is a special case of a possibilistic system, where only occurrence or non-occurrence of a state is recorded. On the other hand, it may be possible to transform any arbitrary system into a selection system as stated above. From the perspective of learning, a selection system is simply a set of positive and negative examples. The occurrence of a state signifies a positive example, and the non-occurrence of a state illustrates a negative example. In order to learn a given concept, during the system reconstruction, we must select, based on some criterion, only those states that contain the attribute relating to that concept. This should enable us to learn a single concept as well as the multiple ones.

Mathematically, we define a nominal system as the following six-tuple [9]

$$B_{no} = (V_{no}, W_{no}, s_{no}, A_{no}, Q_{no}, f_{no}) \quad (5.1)$$

where

- (1)  $V_{no} = \{v_i \mid i \in 1, 2, \dots, n\}$  is a set of variables;
- (2)  $W_{no} = \{V_j \mid j \in \{1, 2, \dots, m\}, m \leq n\}$  is a family of state sets;
- (3)  $s_{no}: V_{no} \rightarrow W_{no}$  is an onto mapping which assigns to each variable in  $V_{no}$ , one state set from  $W_{no}$  ;
- (4)  $A_{no} = s_{no}(v_1) \times s_{no}(v_2) \times \dots \times s_{no}(v_n)$  is a set of all potential aggregate states;
- (5)  $Q_{no}$  is a set of categories or nominal attribute values; and
- (6)  $f_{no}: A_{no} \rightarrow Q_{no}$  is called the nominal system function which represents the information regarding the aggregate states of the system.

We transform the above system to a selection system as follows

$$B_{se} = (V_{se}, W_{se}, s_{se}, A_{se}, Q_{se}, f_{se}) \quad (5.2)$$

where

- (1)  $V_{se} = \{v_i \mid i \in 1, 2, \dots, n+1\}$  is a set of variables, variable  $v_{n+1}$  takes values from  $Q_{no}$ ;
- (2)  $W_{se} = \{V_j \mid j \in \{1, 2, \dots, m\}, m \leq n+1\}$  is a family of state sets;
- (3)  $s_{se}: V_{se} \rightarrow W_{se}$  is an onto mapping which assigns to each variable in  $V_{se}$  , one state set from  $W_{se}$  ;
- (4)  $A_{se} = s_{se}(v_1) \times s_{se}(v_2) \times \dots \times s_{se}(v_{n+1})$  is a set of all potential aggregate states;
- (5)  $Q_{se} = \{0, 1\}$  ; and
- (6)  $f_{se}: A_{se} \rightarrow Q_{se}$  is called the selection function which represents the information regarding the aggregate states of the system.



We define  $f_{se}$  as follows. For all aggregate states  $\alpha_{no}$  and  $\alpha'_{no}$  in  $B_{no}$ ,

$$f_{se}(\alpha_{no} f_{no}(\alpha'_{no})) = \begin{cases} 1 & \text{if } \alpha_{no} = \alpha'_{no} \\ 0 & \text{otherwise} \end{cases} \quad (5.3)$$

Following is an example of the transformed system using (5.3).

$v_1$	$v_2$	$v_3$	$f_{se}$
0	0	flu	1
0	1	flu	0
1	0	flu	0
1	1	flu	0
0	0	cholera	0
0	1	cholera	1
1	0	cholera	0
1	1	cholera	1
0	0	sinus headache	0
0	1	sinus headache	0
1	0	sinus headache	1
1	1	sinus headache	0

It remains to investigate, to what extent, will the new system capture the structure of the original system. Apparently, as there is no loss of information in transforming from one system to another, the transformed system must explain the behavior of the original system. This methodology can be greatly helpful in the domain of expert systems where not only the variables but, in most cases, decisions are also nominal.

## BIBLIOGRAPHY

- [1] D. Angluin and C. Smith, "Inductive inference: Theory and methods", *ACM Comput. Surveys*, 15 (3), 1984, 237-70.
- [2] W. R. Ashby, "Constraint analysis of many-dimensional relations.", *General Systems Yearbook*, 9, 1964, 99-105.
- [3] Y. Bishop, S. E. Fienberg and K. W. Holland, "Discrete Multivariate Analysis: Theory and Practice", Cambridge, MA: MIT, 1975.
- [4] N. M. Blachman, "The amount of information that y gives about X", *IEEE trans. Inform. Theory*, IT-14, Jan. 1968, 27-31.
- [5] R. E. Blahut, "Principles and Practice of Information Theory", Reading MA: Addison-Wiley, 1987.
- [6] D. T. Brown, "A note on approximation to discrete probability distribution.", *Information and Control*, 2 (4), 1959, 386-92.
- [7] R. E. Cavallo and G. J. Klir, "Reconstructability analysis of multi-dimensional relations: a theoretical basis for computer-added determination of acceptable systems models.", *International Journal of General Systems*, 5, 1979, 143-71.
- [8] R. E. Cavallo and G. J. Klir, "Reconstructability analysis: overview and bibliography.", *International Journal of General Systems*, 7, 1981, 1-6.
- [9] R. E. Cavallo and G. J. Klir, "Reconstructability analysis: evaluation of reconstruction hypotheses.", *International Journal of General Systems*, 7, 1981, 7-31.
- [10] R. E. Cavallo and G. J. Klir, "Decision making in Reconstructability analysis.", *International Journal of General Systems*, 8, 1982, 243-55.
- [11] R. E. Cavallo and G. J. Klir, "Reconstruction of possibilistic behavior systems.", *Fuzzy Sets and Systems*, 8, 1982, 175-97.
- [12] J. Cendrowska, "PRISM: an algorithm for inducing modular rules", *International J. Man-Machine Studies*, 27, 1987, 349-70.
- [13] P. Cheeseman, "In defense of probability", in *Proc. Ninth Int. Joint Conf. on Artificial Intelligence*, 2, 1985, 1002-9.
- [14] P. Clark and T. Niblett, "The CN2 induction algorithm", *Machine Learning*, 3, 1989, 261-83.
- [15] P. R. Cohen and E. A. Feigenbaum, "The handbook of artificial intelligence", Vol.3, Los Altos, William Kaufmann, CA, 1982.

- [16] E. Czogala, J. drewniak and W. Pedrycz, "Fuzzy relation equations on a finite set", *Fuzzy Sets and Systems*, 7, 1982, 89-101.
- [17] D. Dubois and H. Prade, "Fuzzy Sets and Systems: Theory and applications", Academic Press, New York, 1980.
- [18] A. O. Duda and P. E. Hart, "Pattern classification and scheme analysis", Wiley, NY, 1973.
- [19] B. R. Gaines and M. L. G. Shaw, "Induction of inference rules for expert systems.", *Fuzzy Sets and Systems*, 18 (3), 1986, 315-28.
- [20] J. G. Ganascia, "Learning with Hilbert cubes", in *Proc. Second European Workshop on Machine Learning (EWSL)*, Bled, Yugoslavia, 1987.
- [21] E. M. Gold, "Language identification in the limit.", *Information Control*, 10, 1967, 447-74.
- [22] I. J. Good, "The estimate of probabilities: An essay on modern bayesian methods", *Research Monograph 30*, MIT, Cambridge, MA, 1965.
- [23] R. M. Goodman and P. Smyth, "An information theoretic model for rule-based expert systems", presented at the 1988 Int. Symp. on Information Theory, 1988, Kobe, Japan.
- [24] A. Hart, "Knowledge Acquisition for Expert Systems", New York: McGraw Hill, 1986.
- [25] D. Haussler, "Bias, version spaces and Valiant's learning framework", *Proc. Fourth Int. Workshop on Machine Learning*, 1987, 324-36.
- [26] M. Higashi and G. J. Klir, "Measures of Uncertainty and Information Based on Possibility Distributions.", *International Journal of General Systems*, 9, 1982, 43-58.
- [27] M. Higashi, G. J. Klir and M. A. Pittarelli, "Reconstruction families of possibilistic structure systems.", *Fuzzy Sets and Systems*, 12, 1984, 37-60.
- [28] E. T. Jaynes, "Prior Probabilities." *IEEE Transactions on Systems Science and Cybernetics*, SSC-4, No. 3, 1968, 227-241.
- [29] E. T. Jaynes, "Where Do We Stand on Maximum Entropy?", In: *The Maximum Entropy Formalism*, edited by R. L. Levine and M. Tribus, MIT Press, Cambridge, Mass., 1979, 15-118.
- [30] P. E. Johnson, "What kind of expert system should a system be?", *J. Med. Philosophy*, 8, 1983, 77-97.

- [31] B. Jones and Oran Brigham, "A note on the determination of a discrete probability distribution from known marginal.", *Information and Control*, 15, 1969.
- [32] B. Jones, "Relating and approximating discrete probability distributions with component distributions.", *Information Sciences*, 6, 1973, 143-50.
- [33] B. Jones, "Determination of reconstruction families.", *International Journal of General Systems*, 8, 1982, 225-8.
- [34] B. Jones, "Recent algorithms in reconstructability analysis.", *SGSR Proceedings*, 1984, 243-45.
- [35] B. Jones, "Determination of unbiased reconstructions.", *International Journal of General Systems*, 10, 1985, 169-76.
- [36] B. Jones, "A greedy algorithm for a generalization of the reconstruction problem.", *International Journal of General Systems*, 11, 1985, 63-8.
- [37] B. Jones, "The cognitive content of system substates.", Special invited paper at the IEEE Workshop on Languages for Automation in Palma de Mallorca, Spain, June 28-29, 1985 and appeared in the *Proceedings of the Symposium on the Cognitive Aspects of Reconstructability Analysis*, Palma de Mallorca, Spain, 1985.
- [38] B. Jones, "Reconstructability analysis for general functions.", *International Journal of General Systems*, 11, 1985, 133-142.
- [39] B. Jones, "Reconstructability considerations with arbitrary data.", *International Journal of General Systems*, 11, 1985, 143-51.
- [40] B. Jones, "K-systems versus classical multivariate systems.", *International Journal of General Systems*, 12, 1986, 1-6.
- [41] B. Jones, "A program for reconstructability analysis.", *International Journal of General Systems*, 15, 1989, 199-205.
- [42] W. Jung, B. Jones and J.Chen, "Optimization of decision trees", *Proceedings of 1991 IEEE Int. Conf. on Tools for AI*, 1991, 522-23.
- [43] G. J. Klir, "Identification of generative structure in empirical data.", *International Journal of General Systems*, 3, 1976, 89-104.
- [44] G. J. Klir and H. J. J. Uyttenhove, "Procedures for generating reconstruction hypotheses in the reconstructability analysis.", *International Journal of General Systems*, 5, 1979, 231-46.
- [45] G. J. Klir, "Architecture of Systems Problem Solving", Plenum Press, New York, 1985.

- [46] G. J. Klir and E. C. Way, "Reconstructability analysis: aims, result and open problems", *Systems Research*, 2 (2), 1985, 141-63.
- [47] G. J. Klir, "Reconstructability analysis: an offspring of Ashby's constraint analysis", *Systems Research*, 3 (4), 1986, 267-71.
- [48] G. J. Klir and B. Parviz, "Relationship between true and estimated possibilistic systems and their reconstructions.", *International Journal of General Systems*, 12, 1986, 319-31.
- [49] S. Kullback, "Information Theory and Statistics", New York, Wiley, 1959.
- [50] P. M. Lewis, "Approximating probability distributions to reduce storage requirements." *Information and Control*, 2 (3), 1959, 214-25.
- [51] F. L. Luconi, T. W. Malone and M. S. S. Morton, "Expert systems: the next challenge for managers", *Sloan Management Review*, Summer 1986, 3-14.
- [52] R. S. Michalski and R. L. Chilausky, "Learning by being told and learning from examples.", *Int. J. Policy Anal. Inform. Syst.*, 4, 1980, 125-61.
- [53] T. M. Mitchell, "Generalization as search", *Artificial Intelligence*, 18 (2), 1982, 203-26.
- [54] M. Pittarelli, "Uncertainty and estimation in reconstructability analysis.", *International Journal of General Systems*, 15, 1989, 1-58.
- [55] M. Pittarelli, "Reconstructability analysis using probability intervals.", *International Journal of General Systems*, 16, 1990, 215-33.
- [56] M. L. Puri and D. Ralescu, "A possibility measure is not a fuzzy measure", *Fuzzy Sets and Systems*, 7, 1982, 311-13.
- [57] J. R. Quinlan, "Induction of decision trees.", *Machine Learning*, 1, 1986, 81-106.
- [58] J. R. Quinlan, "Generating production rules from examples", in *Proc. Tenth Int. Joint Conf. on Artificial Intelligence*, 1987, 304-7.
- [59] J. R. Quinlan and R. L. Rivest, "Inferring decision trees using the minimum description length principle.", *Inform. Comput.*, 80, 1989, 227-48.
- [60] R. L. Rivest, "Learning decision lists", *Machine Learning*, 2, 1987, 229-46.
- [61] E. Sanchez, "Resolution of composite fuzzy relation equations, *Information and Control*, 30, 1976, 38-48.

- [62] C. E. Shannon, "A mathematical theory of communications", Bell Syst. Tech. Journal, 27 (3), July 1948, 379-423.
- [63] J. E. Shore and R. W. Johnson, "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross Entropy", IEEE Transactions on Information Theory, IT-26, No.1, 1980, 26-37.
- [64] P. Smyth and R. M. Goodman, "An information theoretic approach to rule induction from databases.", IEEE Transactions on Knowledge and Data Engineering, 4 (4), 1992, 301-16.
- [65] M. Sugeno, "Fuzzy measures and fuzzy integrals: a survey", in: M. M. Gupta, Ed. Fuzzy Automata and Decision Processes, North-Holland, New York, 1977, 89-102.
- [66] R. E. Valdes-Perez and R. C. Conant, "Information loss due to quantization in reconstructability analysis.", International Journal of General Systems, 9, 1983, 235-47.
- [67] L. G. Valiant, "A theory of learnable.", Communications of ACM, 27 (11), 1984, 1134-42.
- [68] A. K. C. Wong and D. K. Y. Chiu, "Synthesising statistical knowledge from incomplete mixed-mode data", IEEE Trans. Pattern Anal. Mach. Intell., 9, Jan 1987, 796-805.
- [69] R. R. Yager, "A foundation for a theory of possibility, J. Cybernet., 10 (1-3), 1980, 177-204.
- [70] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility", Fuzzy sets and systems, 1, 1978, 3-28.
- [71] L. A. Zadeh, "PRUFA- a meaning representation language for natural languages", International J. of Man-Machine Studies, 10 (4), 395-460.

## VITA

Sudhir K. Trivedi received his B.Sc. degree in Mathematics with Physics and Statistics as optional subjects from Agra University, India in 1979. He received his M.Sc. and Ph.D. degrees in Mathematics from Agra University, India in 1981 and 1985 respectively. He was a research fellow of the Council of Scientific and Industrial Research, India from March 1982 to February 1987. He joined the Ph.D. program in Computer Science at Louisiana State University in the Fall of 1988, where he worked on problems of system reconstruction, design and analysis of related algorithms and learning algorithms based on the process of system reconstruction. Trivedi has been a research assistant in the Robotics Research Laboratory at Louisiana State University. Presently, he is a tenure track assistant professor in Computer Science at Southern University, Baton Rouge.

Besides Systems Theory and Artificial Intelligence, his research interests include Priority Queues, Robotics, Distributed Sensor Integration and Fuzzy Mathematics. Trivedi has published research papers in refereed journals. He is a member of the Association for Computing Machinery, the Society for Industrial and Applied Mathematics, the Indian Mathematical Society, the Operations Research Society of America and the Institute of Management Sciences.

# DOCTORAL EXAMINATION AND DISSERTATION REPORT

**Candidate:** Sudhir Kumar Trivedi

**Major Field:** Computer Science

**Title of Dissertation:** Reconstructability Theory for General Systems  
and Its Application to Automated Rule Learning

**Approved:**

  
Major Professor and Chairman

  
Dean of the Graduate School


**EXAMINING COMMITTEE:**

  
\_\_\_\_\_

  
\_\_\_\_\_

  
\_\_\_\_\_

  
\_\_\_\_\_

  
\_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_

**Date of Examination:** 08-27-93

\_\_\_\_\_

\_\_\_\_\_