بسمه تعالی

جمهوری اسلامی ایران

فرماندهی کل قوا
ستاد کل نیروهای مسلح

دانشگاه و پژوهشگاه عالی دفاع ملی و تحقیقات راهبردی

دفترچه زبان انگلیسی

ویژه مصاحبه دوره دکتری (Ph.D)

رشته مدیریت راهبردی دانش

سال تحصیلی ۱۴۰۲-۱۴۰۱

SYSTEMS METHODOLOGY AND MATHEMATICAL MODELS FOR KNOWLEDGE MANAGEMENT

Yoshiteru NAKAMORI

School of Knowledge Science
Japan Advanced Institute of Science and Technology
Ishikawa 923-1292, Japan
nakamori@jaist.ac.jp

Abstract

This paper first introduces a new discipline knowledge science and the role of systems science in its development. Then, after the discussion on current trend in systems science, the paper proposes a new systems methodology for knowledge management and creation. Finally, the paper discusses mathematical modeling techniques to represent and manage human knowledge that is essentially vague and context-dependent.

Keywords: Service sector, systems engineering, information technology, decision technologies, customer-centric, productivity

1. Introduction

One of the popular ideas among researchers on knowledge management in Japan is as follows: *Knowledge management is not the activity only for resources called knowledge, but the activity to consider how all resources are utilized. You have to make the new structure by repeating new discovery and new creation rather than storing knowledge and considering the combination. Creation can be performed only by people's capability instead of a system.* Apparently this view is from management science. On the other hand, researchers from information science have been trying to establish their own knowledge science using the rapid developing information and communication technology. Can we dream a fusion of these two disciplines?

We understand the limitation of our ability to objectify the real world, the limitation of our ability to understand indirect observation, and the limitation of our ability to analyze things objectively. For this reason, we have been developing a systems methodology that stresses the adaptive learning and stimulation of intuition and creativity of people. The main task is to develop decision or thinking support systems using information technology, which provide system models and system methods with which people can make decisions taking into account social aspects or human relations. But there is a dilemma.

The total system is inseparable, but we cannot perceive well the inseparable whole analytically. Therefore, we usually cut off

ISSN 1004-3756/03/1201/49
CN11-2983/N ©JSSSE 2003

weak links and nonlinear features, and consider individual linear subsystems that we can well imagine. Artificial intelligence inevitably inherits this weakness of human beings. People are usually not satisfied with such pieces of knowledge. We never forget that each element can be included plural subsystems, and that those subsystems are interrelated with each other for this reason. However, it is not easy to develop a logical method to explore the total characteristics based on the study of elements and their relationships.

This paper introduces a new discipline, *knowledge science,* and the role of systems science in its development. Then, after the discussion on current trend in systems science, the paper proposes a new systems methodology for knowledge management and creation, which integrates knowledge in the scientific, creative and human dimensions. The latter half of this paper focuses on the scientific dimension. In the scientific dimension of knowledge creation, one of the most important activities is mathematical modeling based on existing data. Here, the author thinks that the so-called context model is the most promising, which can be converted into a rule-based model often used in the field of artificial intelligence.

Knowledge of people is subjective, ambiguous, and vague in addition to circumstantial. To deal with these properties, the author is developing the ensemble modeling technique, which provides the tendency of people's opinions and at the same time their diversity. This is achieved by extending the traditional multivariate statistical analysis, such as factor analysis or regression analysis, utilizing the fuzzy modeling technique. This ensemble modeling is useful to analyze complex feeling of people in total and to support, for instance, merchandise planning, city planning, or any social decision- makings.

2. Knowledge Science

2.1 Information and Knowledge

Definitions of knowledge range from the practical to the conceptual to the philosophical, and from narrow to broad in scope, which are summarized in (Liebowitz 1999). For instance:

- Knowledge is organized information applicable to problem solving (Woolf 1990);
- Knowledge is information that has been organized and analyzed to make it understandable and applicable to problem solving or decision making (Turban 1992);
- Knowledge is reasoning about information and data to actively enable performance, problem solving, decision-making, learning, and teaching (Beckman 1997).

These definitions require clear distinctions between data, information, and knowledge. Several authors try to distinguish them (Alter 1996; Tobin 1996), also several authors define typologies of knowledge. For instance, Nonaka and Takeuchi (Nonaka and Takeuchi 1995) suggest that the conversion from tacit to explicit knowledge and vice versa is crucial in knowledge creation.

Let us consider the relations between information and knowledge, each of which has the following two meanings:

- Information is (A) knowledge transmitted by character, sign, and voice, etc., or (B) data arranged to be useful for decision-making.

- Knowledge is (C) recognition memorized personally or socially, or (D) judgment or a system of judgment that has objective validity.

Apparently, there is no clear distinction between information and knowledge. However, they are different and each of them is converted to the other.

What is the energy to bring such conversion? Here, let us call it intelligence:

- Intelligence is (E) ability to understand and learn things, or (F) ability to think and understand things instinctively or automatically (e.g. computer intelligence).

As shown in Figure 1, people convert data and knowledge into information for some purpose. They create new knowledge based on data and information. These conversion and creation require existing knowledge and some ability called intelligence.

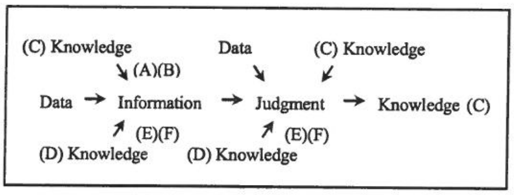

Figure 1 Relationship between Data, Information and Knowledge

Knowledge science is a new discipline related to philosophy, methodology and methods to enhance these abilities. There are mainly two approaches to develop intelligence of human beings: one is from management science and the other is from information science.

2.2 Approach from Management Science

One typical example of the approach from management science is Nonaka theory. Nonaka assumes that knowledge is created through the interaction between tacit and explicit knowledge, and proposes four modes of knowledge conversion as shown in Figure 2 (Nonaka and Takeuchi 1995):

Figure 2 Four Modes of Knowledge Conversion in (Nonaka and Takeuchi 1995)

- Socialization is a process of sharing experiences and thereby creating tacit knowledge such as shared mental models and technical skills.
- Externalization is a process of articulating tacit knowledge into explicit concepts, taking the shapes of metaphors, analogies, concepts, hypotheses, or models.
- Combination is a process of systemizing concepts into a knowledge system. This mode of knowledge conversion involves combining different bodies of explicit knowledge.
- Internalization is a process of embodying explicit knowledge into tacit knowledge. It is closely related to learning by doing.

Corresponding to the above four modes, Nonaka proposes four types of *ba* (roughly means place) (Nonaka et al. 2000):

- Organizing *ba* is defined by individual and face-to-face interactions. It is a place where individuals share experiences, feelings, motions and mental models.
- Dialoguing *ba* is defined by collective and face-to-face interactions. It is the place where individual's mental models and skills are shared, converted into common terms, and articulated as concepts.
- Systemizing *ba* is defined by collective and virtual interactions. This mainly offers a context for the combination of existing explicit knowledge, as explicit knowledge can be relatively easily transmitted to a large number of people in written form.
- Exercising *ba* is defined by individual and virtual interactions. It mainly offers a context for internalization. Here, individuals embody explicit knowledge that is communicated through virtual media, such as written manuals or simulation programs.

Corresponding to the above four modes of knowledge creating process and four types of *ba*, Nonaka proposes to categorize knowledge assets into four types (Nonaka et al. 2000):

- Experimental knowledge asset consists of the shared tacit knowledge that is built through shared hands-on experience amongst the members of the organization, and between the members of the organization and its customers, suppliers and affiliated firms.
- Conceptual knowledge asset consists of explicit knowledge articulated through images, symbols and languages. They are the assets based on the concepts held by customers and members of the organization.
- Systemic knowledge asset consists of systematized and packaged explicit knowledge, such as explicitly stated technologies, product specifications, manuals, and documented and packaged information about customers and suppliers.
- Routine knowledge asset consists of the tacit knowledge that is routinized and embedded in the actions and practices of the organization. Know-how, organizational culture and organizational routines for carrying out the day-to-day business of the organization are examples of routine knowledge assets.

2.3 Approach from Information Science

Approach from information science is as follows. Information engineering relates not only to the field of life science, such as psychology and physiology, but to sensitivity, intellect, thought, etc., and is based also on the field of social and cultural science, such as economy, politics, and culture. Although this is commonly recognized, information engineering has been developed as a study of computer hardware and software, and its application, i. e, computer science.

Information originally accompanies all the sides of human activities. However, since it is hard to feel directly, unlike substance or energy, its conceptualization was behind. Although the appearance of the computer contributed to its conceptualization greatly, it limited the range of information as the object of technology to the computer and its circumference. It is necessary to expand this range and to bring close to the system of information over life science, social science, and cultural sciences.

In order to systematize the information technology covering a wide area, it is most intelligible to focus on a fundamental concept and to position other concepts in connection with that. Let the basic concept be knowledge. Although knowledge is deeply concerned with intelligence, the knowledge science treating the knowledge expressed with the sign or the non-sign develop as a different thing from the conventional information technology, and is also beginning to obtain an effect. It is an important subject to systematize this technology.

From this perspective, every method in artificial intelligence can be considered a method in knowledge science. Figure 3 shows the hierarchy of knowledge science from the viewpoint of information technology.

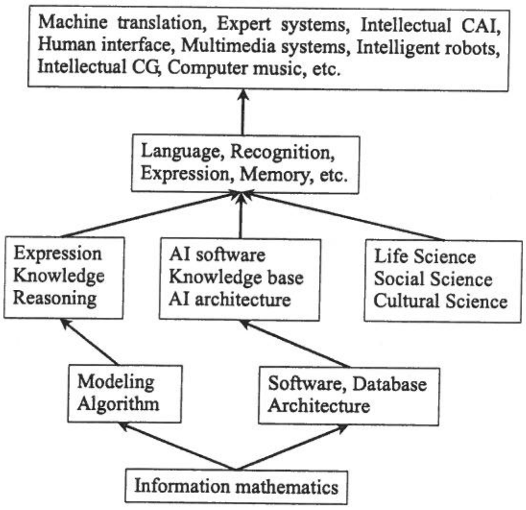

Figure 3 Hierarchy of Knowledge Science from the Viewpoint of Information Technology

Recall that the second definition of intelligence, which is the ability to think and understand things instinctively or automatically. Knowledge science includes the following topics from information science:

- Knowledge creating methodology: hierarchical conceptual modeling for problem solving, which is a process of forming and creating conceptual knowledge and plays a fundamental role in information processing.
- Knowledge-based systems: research on the theory and design methodology of evolving knowledge systems, particularly networked information systems and human brain neural networks, which are studied from analytical and synthetical points of view.
- Knowledge structure: study of knowledge structure in terms of concept making based on the hierarchization, abstraction, and instantiation of concepts, and theory and methodology on the representation, acquisition, and utilization of knowledge.
- Creativity support systems: principles for building knowledge-based systems and knowledge creating environment, and application of the principles to the real world intellectual problems.

3. Systems Methodology

3.1 Pluralism or Fusion

To integrate the above approaches and establish a new discipline is a quite natural idea, and then School of Knowledge Science was established in Japan Advanced Institute of Science and Technology in 1998. This is, however, not the first trial in our history, and most of them are not necessarily successful. Something is necessary for the success, which may be the idea of system. Systems science may have an important role for the success of establishing knowledge science. However, there are also two different schools in the field of systems science: hard and soft schools, which roughly correspond to the fields of information science and management science, respectively. Something is necessary more.

Most systems thinkers today might agree the methodological pluralism or complementarism (Jackson 1999) discussed in the fields of sociology and organizational studies. The pluralism means that it is necessary to develop and employ a wide range of heterogeneous methodological devices so as to investigate and deal with the complexities. The complementarism suggests that those devices can communicate with and support each other as some methods are good at tackling hard problems while others at soft issues given that human problems are conditioned and constituted by both hard problems and soft issues (Zhu 1999).

There is a difference between considerations of systems thinkers in UK and China. The researchers of UK often use the term pluralism or complementary use of different methodologies or methods. On the other hand, Chinese systems thinkers often use the term integration or fusion of different approaches. We see the difference between Westerners and Easterners in the difference between complementary use and fusion. But, this is not always true. Maybe, we should say that there are differences in the process of complementary use of several approaches depending on the background of researchers. Systems methodology to deal with knowledge should be a fusion of different methodologies rather than just a set of complementary methodologies.

The most reliable knowledge source is the scientific investigation that produces public knowledge. This is objective, unique, universal, and repeatable. On the other hand, knowledge obtained in social science includes meanings given by people inevitably, which are wisdom-based knowledge, insight-based knowledge, and experienced-based knowledge. These kinds of knowledge are subjective, vague, ambiguous, and circumstantial. Information science treats the former type of knowledge, while management science has to deal with the latter type of knowledge. This is the main reason that the effort of establishing interdisciplinary school has not been successful. Knowledge science has to treat different kinds and levels of knowledge.

3.2 Knowledge Creating System

This section introduces a systems methodology for integration, management and creation of different types of knowledge. This is a research in knowledge science that creates justified true belief, or systemic knowledge. This study uses approaches in social and natural sciences complementarily. The first one is scientific approach that uses physical laws, data analysis, etc. The second is information science, especially a large-scale computer simulation and the networking technology. The third is a method in social science, which is related to forming partnerships among social members.

The developing system can be called a knowledge creating system. The system integrates statistical data and individual persons' fragmentary knowledge, and then creates new knowledge nobody had before. Such knowledge must be tacit, otherwise someone including the system had it; this is a contradiction. Therefore, the system should have a process to convert tacit knowledge into explicit knowledge. This means that the members of the project or relevant people constitute a part of the system. This systems methodology itself is a system consisting of five subsystems as shown in Figure 4.

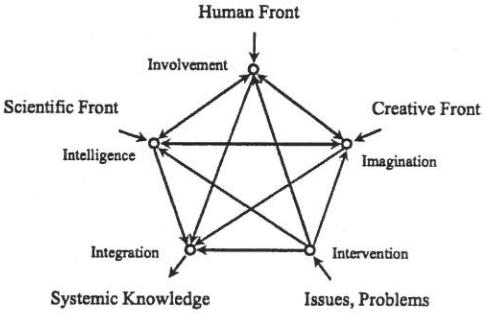

Figure 4 A knowledge Creating System (*i*-System)

1) Intelligence: We take action in a situation that we were not originally involved in. We here consider what kinds of knowledge are necessary to solve the faced problem, and request the following three subsystems to collect them. Here knowledge is a problem.

2) Imagination: We enhance our ability to understand and learn things. We here collect necessary data and information, analyze them with a scientific attitude, and make a model for simulation or optimization. Here, knowledge is a model.

3) Involvement: We build our own idea on new or exciting things. Here, we simulate complex phenomena based on partial knowledge, using information technology. Here, knowledge is scenarios.

4) Integration: We raise our and other people's concern and enthusiasm. We hear opinions of people by organizing a meeting or questionnaire survey. Here, knowledge is opinions.

5) Intervention: We combine different knowledge so that they are closely linked. Here, we evaluate reliability and justifiability of outputs from three subsystems, and integrate them. Here, knowledge is solutions.

We evaluate the knowledge-creating system from the following viewpoint:

- Are the system, actors and contents well defined?
- Is its foresight power enough?
- Is the totality achieved?
- And, is it actually useful?

An example of complementary use of people and computer is as follows: Suppose now we have a problem of how to activate ecological industry. At *Intelligence*, we will make a model with computer based on ideas of people. At *Imagination*, we carry out computer simulation with assumptions given by people. At *Involvement*, we develop a network with the help of information technology initiated by people. Then, at *Integration*, we will build a strategic scenario-based system by consulting the relevant people. See Figure 5.

This methodology is a system because it has the following properties:

- hierarchical structure,
- emergent characteristics,
- function of communication, and
- function of control and feedback.

For example, if we consider sustainable development, the role of subsystem *Intelligence* is prediction based on scientific knowledge. To achieve this task, this subsystem asks the lower system to develop a mathematical model, and then the subsystem *Involvement* of this lower system will collect necessary data consulting the relevant people. Then, the next lower system will start collecting numerical data, qualitative data, and scenario data. The subsystem *Imagination* of this level will ask the next lower system to collect assumptions and ideas with their possibilities. In such a way, this knowledge-creating system has a hierarchical structure with functions of communication and control. We can define the integrated knowledge as emergence. See Figure 6.

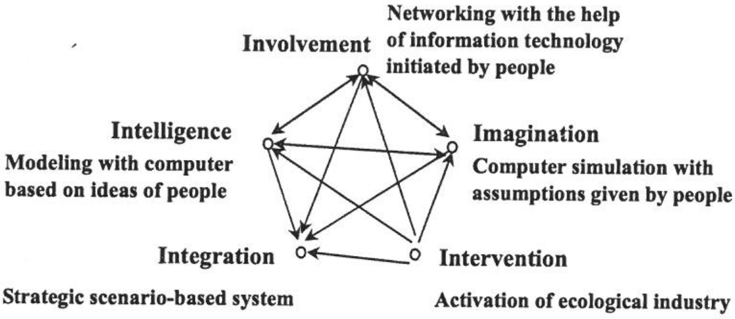

Figure 5 Utilization of Capabilities of People and Computers Complementarily

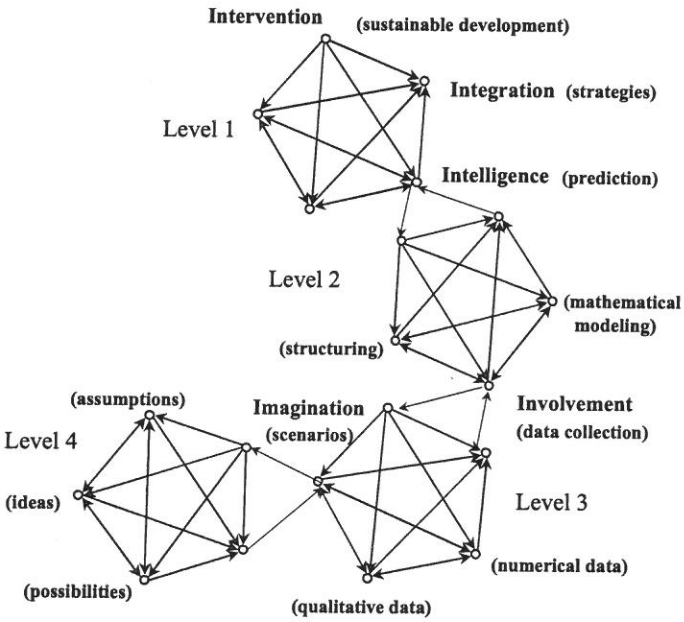

Figure 6 The Methodology Has the Hierarchical Structure, Emergent Characteristics, Function of Communication, Function of Control and Feedback

How is the trans-disciplinary knowledge exchange achieved? The subsystem *Intelligence* is mainly based on natural sciences, mathematics, and engineering. The subsystem *Imagination* is related to information science, economics, and statistics. The subsystem *Involvement* is mainly based on management science, social science, and cultural sciences. The other two subsystems *Intervention* and *Integration* are of course related to systems science. Different disciplines are used to determine boundary conditions or to check consistency of knowledge of subsystems. However, this is not performed automatically. This is an interactive system. See Figure 7.

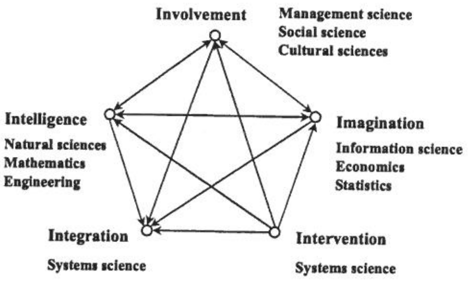

Figure 7 Trans-Disciplinary Knowledge Exchange

3.3 Illustrative Example

For instance, consider the problem of writing scenarios of sustainable development. The roles of subsystems are determined as follows:

Level1 (MAIN)

- *Intervention:* Collect the following knowledge for writing scenarios.
- *Intelligence:* Collect the knowledge on the relation between economy and environment.
- *Imagination:* Collect the knowledge on innovation in ecological industry.
- *Involvement:* Collect the knowledge on construction of resources circulation society.
- *Integration:* Write scenarios by integration of the above knowledge.

Subsystems *Intelligence, Imagination,* and *Involvement* correspond to scientific front, creative front, and human front, respectively. However, each assigned task cannot be carried out only by one discipline. That is, each task usually requires knowledge from other disciplines. Therefore, each subsystem has the same structure as Level1(MAIN) as follows:

Level2 (Subsystem=*Intelligence*)

- *Intervention:* Collect the knowledge related to the relation between economy and environment.
- *Intelligence:* Collect the knowledge on models to predict future environment based on socioeconomic conditions.
- *Imagination:* Collect the knowledge on stories of socioeconomic development.
- *Involvement:* Collect the knowledge on confidence of prediction models and stories.
- *Integration:* Construct a knowledge model to predict future environment based on the scenarios of economic development.

Level2 (Subsystem=*Imagination*)

- *Intervention:* Construct models to estimate innovation of technology and its effect with a long-term viewpoint.
- *Intelligence:* Collect the scientific knowledge on the present condition and a view of environmental technology.
- *Imagination:* Collect the knowledge on the possibility of technical innovation.

- *Involvement:* Collect the sociological knowledge on creation of ecological industry.
- *Integration:* Construct a knowledge model to predict environmental improvement effects from technical innovation scenarios.

Level2 (Subsystem=*Involvement*)

- *Intervention:* Construct models of resources circulation and the energy-minimum- type society.
- *Intelligence:* Collect the knowledge on the present condition and problems of social systems.
- *Imagination:* Collect the knowledge on the systems, policies, and international cooperation.
- *Involvement:* Collect the knowledge on the role of international organizations, the governments, and local governments.
- *Integration:* Construct a knowledge model to predict future environment from social change scenarios.

The task of subsystem *Intelligence* in Level 2 (Subsystem=*Intelligence*) is further divided into five tasks with different viewpoints. For instance:

Level3 (Subsystem=*Intelligence*)

- *Intervention:* Collect the knowledge on models to predict future environment based on socio- economic conditions.
- *Intelligence:* Construct mathematical models.
- *Imagination:* Consider development stories.
- *Involvement:* Collect necessary data.
- *Integration:* Construct a knowledge model.

The task of subsystem *Involvement* in Level 2 (Subsystem=*Involvement*) is further divided into five tasks of different viewpoints. For instance:

Level3 (Subsystem=*Involvement*)

- *Intervention:* Collect the necessary data.
- *Intelligence:* Collect numerical data.
- *Imagination:* Collect text data.
- *Involvement:* Collect scenario data.
- *Integration:* Construct a database.

The task of subsystem *Imagination* in Level 2 (Subsystem=*Imagination*) is further divided into five tasks with different viewpoints. For instance:

Level3 (Subsystem=*Imagination*)

- *Intervention:* Collect scenario data.
- *Intelligence:* Consider possibilities.
- *Imagination:* Collect ideas.
- *Involvement:* Collect assumptions.
- *Integration:* Construct a scenario base.

3.4 Applications

As applications of the knowledge-creating system, two ongoing projects are introduced in this section.

1) Environment System Model: Figure 8 shows an environment system model. The main model is called the environment framework model that integrates environmental knowledge and carries out a short-term environmental prediction. To this end, we have to develop a new systems methodology that emphasizes knowledge creation, a new simulation technique based on the complex adaptive systems methodology, and complex systems modeling techniques using the soft computing methods.

This is an interdisciplinary research that integrates knowledge from different fields. Interaction between people and computers are emphasized to obtain agreements. We are employed in developing a system model for Ishikawa prefecture, Japan.

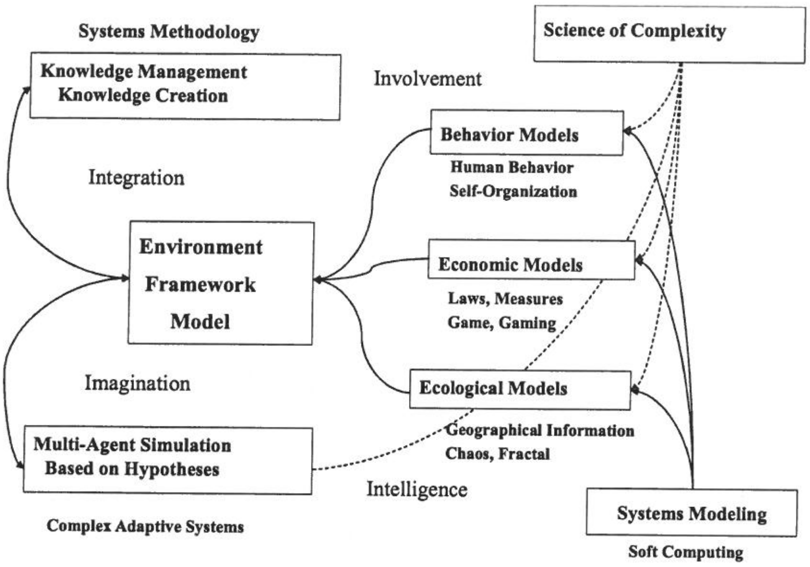

Figure 8 Development of an Environment System Model

- Scientific Front: we have gathered hard data, developed a part of the framework model, and we are developing the rest part of the model using soft modeling techniques.
- Human Front: we have investigated environmental activities using the computer network, and we have gathered soft data by a questionnaire survey.
- Creative Front: we have constructed a platform to carry out agent-based simulations, and we are designing a model of ecological product and consumption.

2) Ecological Business Model: We are developing an ecological business model in cooperation with the local government. Recently, the ecological business attracts attention from two aspects in Japan. One is the possibility of new business in the circumstances of long economic depression. Another is the possibility of promoting people's attitude or behavior to the environment. The aim of this research is to design an ecological business model that reflects the needs of consumers and provides idea of products or services with a less load on the environment.

An important tool we have to develop is a public hearing system as shown in Figure 9. The tasks of this system are to collect soft data from both consumers and enterprises, to provide a site for a public auction of ecological products. The idea exchange system offers a place to create new knowledge with the cooperation of enterprises.

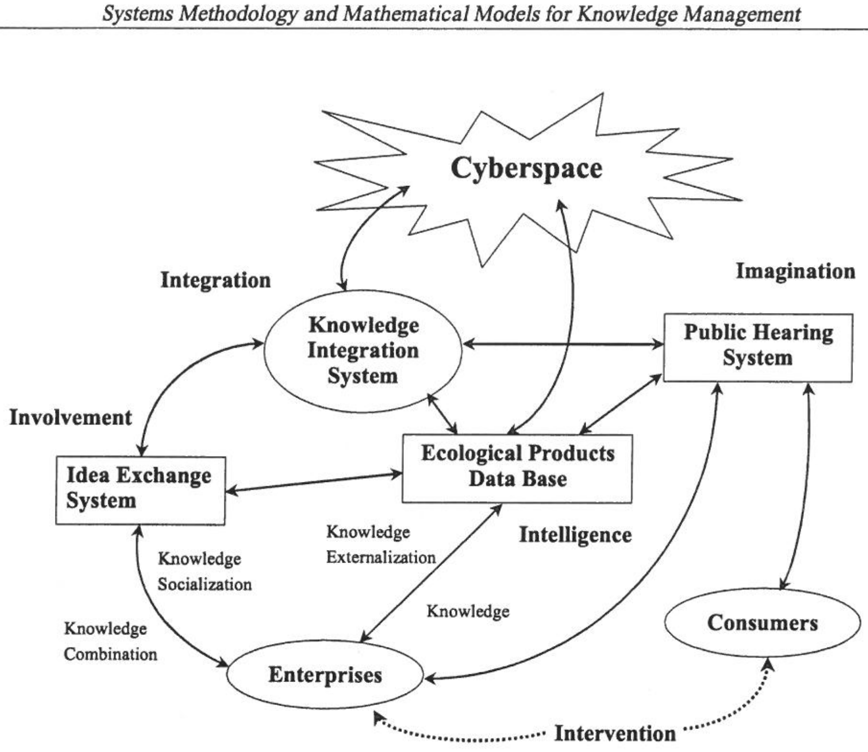

Figure 9 Development of an Ecological Business Model

4. Mathematical Knowledge Models

Here, we discuss mathematical modeling techniques to represent human knowledge that is essentially vague and context dependent.

An example is shown in Figure 10. The vertical axis is the level of water pollution, for instance, biochemical oxygen demand, and the horizontal axis is the spread level of the sewer system. The data of respective area moves like as respective circles or rectangles. The data indicated by white and black circles came from rapidly developed area, while the data indicated by white and black rectangles came from gradually developed or undeveloped area.

This means that we have to develop different models depending on the conditions. Also, we have to consider the individuality. That is, each area has its own history of development. From this observation, the authors are developing techniques to build context-dependent ensemble models, which may be the most promising models for knowledge representation because knowledge is context-dependent, circumstantial, vague and ambiguous.

4.1 Context Models

In (Kruse et al. 1993), the authors introduce the context model as an approach to representation, interpretation, and analysis of imperfect data. They think that the origin of imperfect data is due to situations, where either we are not able to specify an object or we only need to use linguistic descriptions about the object, which are more natural and understandable for human beings. In addition, the specific meaning of vague concepts in human thinking and communication is always determined by context, by personal view, etc., i.e., their interpretation depends on which context they are uttered in.

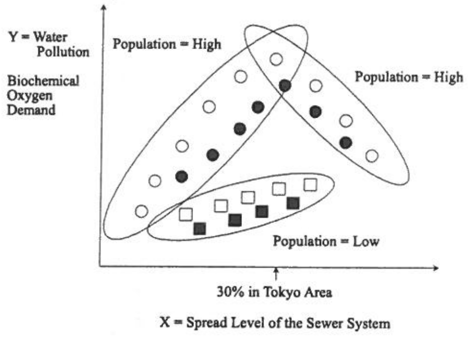

Figure 10 Context Dependency and Individuality

A context model is defined as a triple $< D, C, A_C(D) >$, where D is a nonempty universe of discourse, C is a nonempty finite set of contexts, and the set $A_C(D)$ is the set of all vague characteristics of D with respect to C, i. e, the set of all mappings from C to 2^D. For instance, consider the height of adults: D= [1.0m, 3.0m], C ={Japanese, American,⋯}. Then, if a = tall, we guess tall(Japanese) =[1.7m, 1.8m], and tall(American)=[1.8m, 1.9m].

If there is a finite measure P_C on the measurable space $(C, 2^C)$, then $a \in A_C(D)$ is called a valued vague characteristic of D with respect to P_C. Then, a quadruple $< D, C, A_C(D), P_C >$ is called a valued context model. By this context model, each linguistic term $a \in A_C(D)$ may be semantically represented by the fuzzy set A as follows:

$$\mu_A(x) = \sum_{c \in C} P_C(c)\chi_{a(c)}(x) \qquad (1)$$

where $\chi_{a(c)}$ is the characteristic function of $a(c)$. Intuitively, while each subset $a(c)$, for $c \in C$, represents the c's view of the vague concept a, the fuzzy set A is the result of a weighted combinational view of the vague concept.

4.2 Ensemble Models

Let us now consider the varieties of knowledge or feeling of people for an object that could be either a concrete reality or an abstract concept. A set of qualitative data obtained by subjective evaluation for an object usually has a large variance reflecting tastes and preferences of individuals. This is related, in a sense, to the principle of incompatibility claimed in Zadeh (Zadeh 1973). It is sensible to express such fluctuations by fuzzy numbers to treat vagueness and uncertainty of the feeling of individuals.

The fuzzy-sets theory (Zadeh 1965) has been applied to multivariate data analysis, and nowadays, fuzzy data analysis is widely used in various fields. The possibilistic regression analysis (Tanaka 1987) is one of the most successful applications of fuzzy concept and fuzzy logic. There are also many research activities to apply the fuzzy-sets theory to other areas in multivariate data analysis. However, there is still room for consideration in defining fuzzy sets based on the given data set and in applying the extension principle (Zadeh 1975) to modeling of aggregating functions. In multivariate data analysis, the squares of original variables often have an essential role. Operation of fuzzy numbers sometimes leads a result incompatible with our intuition when membership functions are defined on the original variable space.

We have developed an ensemble modeling

technique (Nakamori and Ryoke 2001). First, an average multivariate model is identified, and then by a data mapping technique, the personal data is mapped into the parameter space of the multivariate model. This mapping preserves, in some sense, the relative positions of personal perception of people. After data mapping, membership functions are determined in the parameter space to develop a fuzzy multivariate model. A notable fact of this approach is that it expresses personal positions in the parameter space of a multivariate model, preserving the relative personal positions in the original data space as much as possible. The idea is applied to factor analysis and quantification method. The latter corresponds to regression analysis for categorical data. The ensemble linear modeling technique is introduced in the following.

4.3 Eensemble Linear Models

This section introduces an ensemble linear modeling technique. Suppose that the data set of objective variable is given by

$$\{y_{mk}\},\ m=1,2,\cdots,M;\ k=1,2,\cdots,K \quad (2)$$

Here, m indicates an object for evaluation. For instance, imagine that a lady is about to purchase a new dress from among M dresses, or a personnel manager is interviewing M candidates for employment. Evaluators are indicated by $k=1,2,\cdots,K$. They are often called subjects. Attributes of objects are indicated by $i=1,2,\cdots,I$. Denote the value of an attribute i of the object m by x_{mi}. We here treat the case that there are K outputs $y_{m1}, y_{m2},\dots, y_{mK}$ for the same input vector $(x_{m1}, x_{m2},\dots, x_{mI})^{\mathrm{T}}$.

1) *Regression Analysis with Average Data:* Based on the average data in evaluators:

$$y_m = \frac{1}{K}\sum_{k=1}^{K} y_{mk},\ m=1,2,\cdots,M \quad (3)$$

a regression model is first identified:

$$y_m = a_0 + \sum_{i=1}^{I} a_i x_{mi} + e_m,\ m=1,2,\cdots,M \quad (4)$$

Let us introduce the following notation:

$$y=(y_1, y_2,\dots, y_M)^{\mathrm{T}} \quad (5)$$

$$X=\begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1I} \\ 1 & x_{21} & x_{22} & \cdots & x_{2I} \\ & & \vdots & & \\ 1 & x_{M1} & x_{M2} & \cdots & x_{MI} \end{pmatrix} \quad (6)$$

$$a=(a_0, a_1,\dots, a_I)^{\mathrm{T}} \quad (7)$$

$$e=(e_1, e_2,\dots, e_M)^{\mathrm{T}} \quad (8)$$

If $X^T X$ is non-singular, by the method of least squares:

$$\|e\|^2 \rightarrow \text{minimize} \quad (9)$$

the regression coefficients and residuals are given as

$$\hat{a}=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y \quad (10)$$

$$\hat{e}=y-X\hat{a} \quad (11)$$

2) *Data Mapping:* Let us map the values of output by each evaluator:

$$y_k=(y_{1k}, y_{2k},\dots, y_{Mk})^{\mathrm{T}} \quad (12)$$

into the parameter space by the equation:

$$\hat{a}_k=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(y_k-\hat{e}) \quad (13)$$

The following facts are easily verified:

- ♦ The average of regression parameters of all evaluators is equal to the parameter calculated by the average data:

$$\frac{1}{K}\sum_{k=1}^{K}\hat{a}_k = \hat{a} \qquad (14)$$

- The variance-covariance structure between evaluators in the output data is preserved in the parameter space in the following sense:

$$\|y_k - y\|^2_{X(X^TX)^{-1}X^T} = \|\hat{a}_k - \hat{a}\|^2_{X^TX} \qquad (15)$$

It should be noted that the map (13) has the following implication: $\hat{a}_k$ minimizes the square norm $\| y_k - X\hat{a}_k - \hat{e}\|^2$. That is, $X\hat{a}_k + \hat{e}$ gives the least square error when estimating y_k by the form $Xa_k + \hat{e}$.

3) New Fuzzy Linear Models: We can now introduce a fuzzy vector A in the parameter space using the regression parameters $\hat{a}, \hat{a}_1, \cdots, \hat{a}_k$.

Let us define the membership function of A by:

$$\mu_A(a) = \exp\{-\| a - \hat{a}\|^2_{D_A^{-1}}\} \qquad (16)$$

Then the output membership function of the fuzzy model:

$$Y = Ax \qquad (17)$$

is given by

$$\begin{aligned}\mu_Y(y) &= \max_{\{a|y=a^Tx\}} \mu_A(a) \\ &= \exp\{-(y-\hat{a}^Tx)^2(x^TD_Ax)^{-1}\}\end{aligned} \qquad (18)$$

In Tanaka's method (Tanaka and Ishibuchi 1995), the positive-definite matrix D_A is obtained by introducing a constraint for each data y_m, and solving a linear programming problem. However, in the present case, it can be defined as follows:

- Let the variance-covariance matrix of parameters $\hat{a}, \hat{a}_1, \cdots, \hat{a}_k$ in the *(I+1)*-dimensional space be S_A, and let

$$D_A = cS_A \qquad (19)$$

- Determine c in the above equation so that the inequality

$$\mu_A(a_i) \geq h \qquad (20)$$

holds for all i. Here, h should be determined subjectively.

4) *Numerical Example:* Table 1 shows the case in which five persons give different output values $y_{mk}\ (m = 1,2,\cdots,10; k = 1,2,\cdots,5)$ for the same input $x_m\ (m = 1,2,\cdots,10)$.

Table 1 Input Data and Plural Output Data

m	1	2	3	4	5
x_m	2	4	6	9	12
y_m	4	7	6	8	7
y_{m1}	2.0	8.0	4.0	9.0	6.5
y_{m2}	5.5	4.5	6.0	10.0	3.5
y_{m3}	5.0	6.5	7.0	6.0	8.5
y_{m4}	4.5	8.5	3.5	6.5	9.0
y_{m5}	3.0	7.5	4.5	8.5	7.5
m	6	7	8	9	10
x_m	13	14	16	19	20
y_m	9	12	9	14	10
y_{m1}	10.0	11.0	12.0	14.5	12.5
y_{m2}	11.5	14.5	11.0	15.5	10.5
y_{m3}	7.5	12.5	7.5	15.0	9.0
y_{m4}	9.5	10.5	8.0	13.0	8.5
y_{m5}	6.5	11.5	6.5	12.0	9.5

The regression coefficient vector by the average data is calculated as

$$\hat{a} = (3.7885, 0.4097)^T \qquad (21)$$

The individual regression coefficient vectors computed by Equation(13) are given as follows:

$$\hat{a}_1 = (2.4908,\ 0.5617)^T \quad (22)$$
$$\hat{a}_2 = (3.6351, 0.4883)^T \quad (23)$$
$$\hat{a}_3 = (4.4056, 0.3517)^T \quad (24)$$
$$\hat{a}_4 = (4.4771, 0.3194)^T \quad (25)$$
$$\hat{a}_5 = (3.9342, 0.3275)^T \quad (26)$$

Figure 11 shows the data of five persons and regression lines, where the black dots indicate the average values. The line indicated $\hat{y}$ corresponds to the regression line obtained by the average data. Incidentally, the lines by $y_1, y_2, \cdots, y_5$ are very close to those calculated by the method of least squares independently.

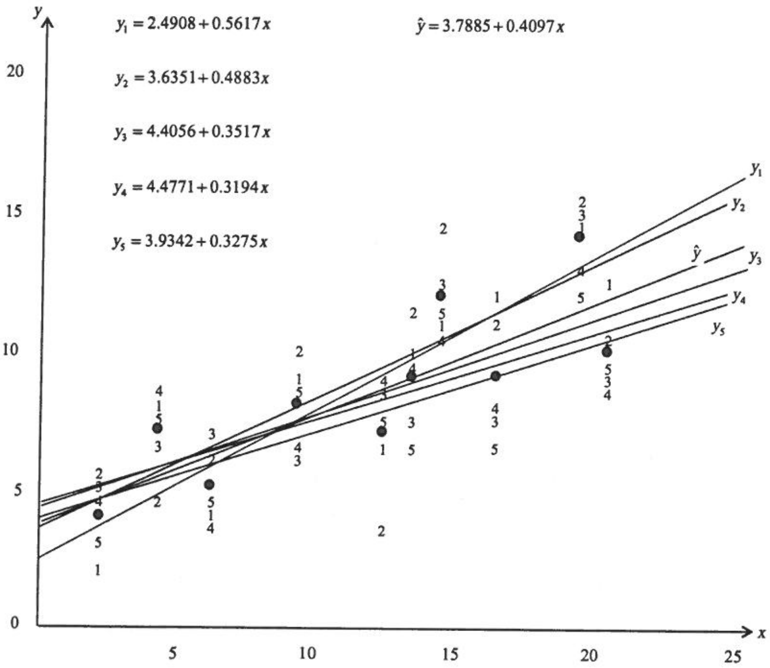

Figure 11 Data of Five Persons and Regression Lines by the Proposed Method

If we define D_A as the quintuple of the variance-covariance of regression coefficients of five persons, then all regression coefficients have the possibility greater than 0.5 as shown in Figure 12. In this case D_A is given by

$$D_A = \begin{pmatrix} 2.5839 & -0.3192 \\ -0.3192 & 0.0476 \end{pmatrix} \quad (27)$$

There is a possibility that the variance-covariance matrix becomes singular. For such a case, it should be regularized by, for instance, adding a small positive number to the diagonal elements. Anyhow, the modeler's judgment is required to determine D_A.

Figure 13 shows the data of five persons and the output membership function (the proposed fuzzy regression model).

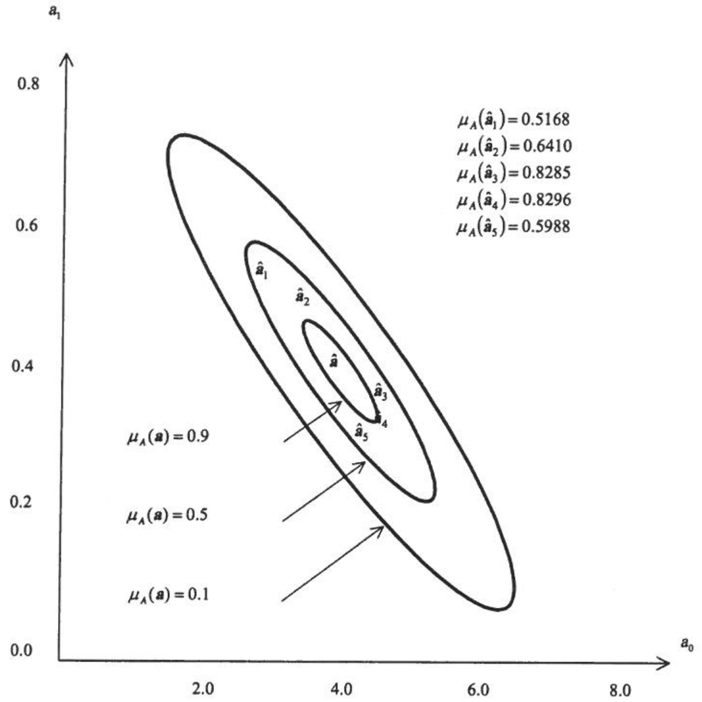

Figure 12 The Membership Function of Regression Coefficients

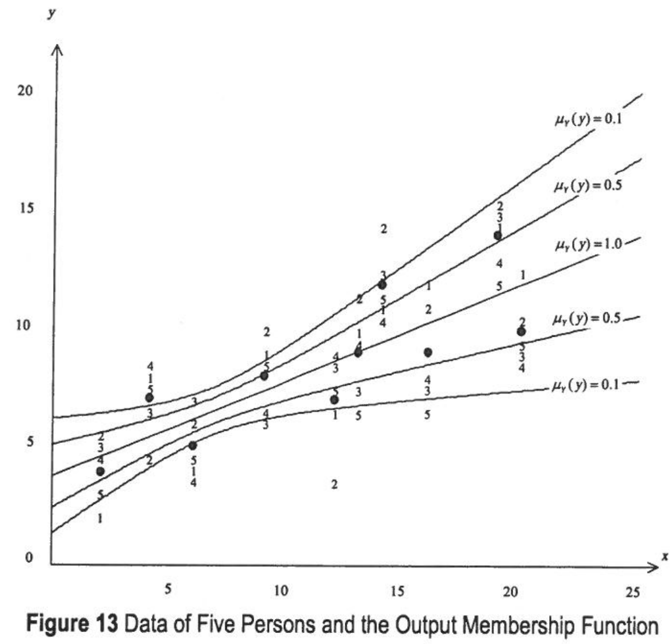

Figure 13 Data of Five Persons and the Output Membership Function