Predictive modeling of everyday behavior from large-scale data

— Learning and inference from Bayesian networks based on actual services —

Yoichi Motomura

[Translation from *Synthesiology*, Vol.2, No.1, p.1-11 (2009)]

Daily life behavior modeling is discussed. This modeling framework consists of statistical learning, probabilistic reasoning, user modeling, and large-scale data collecting technologies. Bayesian networks can represent causality relationship as graphical structures. Such models should include situations and contexts of daily life behavior through real services. In order to collect large-scale data connected with them, we have to provide real services supported by many users. This concept is named "Research as a service" and discussed in this paper.

Keywords **:** Bayesian network, statistical learning, probabilistic reasoning, user model, behavior analysis, knowledge circulation

1 Introduction

The range of applications of information processing technology is steadily increasing. At the same time, information services to aid in everyday life are increasingly in demand. Therefore, a model is necessary which describes the activities of daily life in a quantifiable way, in terms of what a person is trying to accomplish in various circumstances. Using such a quantitative theoretical model, we consider a system that predicts background requirements and expected results from the user's activities, rapidly implements them, and makes possible the development of new services that aid activities of everyday life. Additionally, by continuously implementing such cooperative operations with people during everyday life, it becomes possible to acquire meaningful data in large quantities not previously obtainable in a laboratory environment. Using this large-scale data, it is possible to bring about a cycle in which services continue to be used while the model is constantly being updated.

However, during these daily activities, information processing based on uncertain information (such as predictions that result in indeterminate or incomplete observational information) is of fundamental importance. What is needed is a paradigm shift from the deterministic approach, which has until now played a central role in system recording methods, to a non-deterministic approach. The non-deterministic approach is an approach to calculation in which ambiguous or uncertain information is processed *as is*, as far as possible. Calculations are made with the probability distribution as an object variable, along with the stochastic inference, which makes the prediction [1]. This stochastic inference has come to be used naturally as a naïve Bayes model or a Hidden Markov Model (HHM) using, for example, a pattern recognition device that maximizes the posterior probability. Further, in order to control the system based on decision theory, express useful knowledge, and perform complex processing, calculations with high-dimensional probability distributions involving multiple variables are necessary. As the number of variables becomes enormous, calculations involving high-dimensional probability distributions become complex; therefore, one has no recourse but to approximate locally using low-dimensional probability distributions. In order to facilitate this, a graph structure is introduced which stipulates the relationship between variables. As a multidimensional probability distribution model having this type of graph structure, we have the example of a Bayesian network [2]. Bayesian networks are general models that stipulate dependencies among many variables by a conditional probability and network structure. Bayesian networks can construct a model by statistical learning from large-scale data, which in turn becomes an important feature in handling uncertainty.

In the current work, after discussing the non-deterministic approach and probabilistic modeling, together with Bayesian networks and techniques of constructing models that use them for predicting human activities in everyday life, actual cases in which they are applied are discussed. Finally, a hypothesis about "Research as a Service," the construction of which has become inevitable in the process of implementation, is proposed and discussed.

2 Selection by Non-deterministic Approach

In real-world problems, we want to know the situation (value) or the possibility (probability) of objects that cannot be observed directly (latent variables). This type of uncertainty inevitably enters into computing when humans are considered as the object. When a system implements any task, the tasks are modeled within the system and considered to be the object

Center for Service Research/Digital Human Research Center, AIST Tsukuba Central 2, Umezono 1-1-1, Tsukuba 305-8568, Japan/Aomi 2-41-6, koto-ku 135-0064, Japan E-mail : y.motomura@aist.go.jp

Received original manuscript September 24, 2008, Revisions received January 13, 2009, Accepted January 13, 2009

of calculation operations. In other words, the program can be understood to be the model and computing operations of the object tasks, coded by means of a programming language. Further, the same operation cannot be performed for all users; rather, it is necessary to invoke both a model of the task and a model of the user within the system. Ordinarily, models of the task (process) can often be given clear descriptions; however, for users, in order to treat uncertainties associated with humans, a non-deterministic analytical model is often necessary. In the current state of affairs, clearly modeling the latent variables of human relations, such as the intentions or requests of users, is difficult; and one is compelled, in the first place, to describe this type of situation using a non-deterministic framework. In addition, when a variety of users utilize the system in various circumstances, stipulating beforehand all of the most appropriate operations the system should take is a difficult problem. The system designer should design in advance the capabilities offered by the system; however, answers to questions such as what the system user requests, how the user will react to information or services offered, were the system operations correct, were the user expectations met, etc., will not be known until the system is executed or even after it is executed. In other words, it is difficult to decide, in advance, the most appropriate design of operations for users. Consequently, merely using a non-deterministic framework to operate the system as requested by, or as expected by, users is inadequate; and having predicted user reaction at execution time, frameworks allowing dynamic construction of user models become very important in optimizing these reactions and evaluations. This is the uncertainty associated with humans.

Information to be calculated appears in large amounts, and large gaps emerge between computable amounts. However, handling uncertainty is necessary. For example, through the spread of the Internet, we find that limitations exist; and handling it directly necessitates facing an unwieldy amount of data. It is possible to calculate the frequency with which a given web page is read by all users; however, deterministic processing, such as counting this for all web pages, is not realistic. In this case, Google's PageRank is calculated by modeling the transition probability among web pages non-deterministically as a stationary stochastic process [3]. In other words, we describe deterministically the construction of source pages or related links, and although this is a computer-based or deterministic method, it is not contained within a deterministic framework. Being a strategy that uses a non-deterministic model, it can respond to an explosion in description quality or the number of data points. Coping with uncertainty in a system involving this type of real world or large-scale data that includes humans is highly desirable in an artificial intelligence system, in order to tackle real societal problems; and here, describing problems using a non-deterministic model is one solution.

Even if problems involving uncertainty are described by a non-deterministic computational model, the current deterministic computer processing brings to mind the computational theory of Marr [4] in which the following types of questions are considered independently: what is to be computed, how does the calculation method write the computational process (algorithm level), how is it to be implemented, etc. In other words, even if implemented by a program described by a deterministic computer language on a deterministic silicon chip computer, and, as in the previous web example, even if the original data or mechanism is deterministic, it can be profitable to think of the model as being calculated non-deterministically. As far as "toy problems" are concerned, it is sufficient to consider calculated quantities deterministically; however, one cannot avoid using a description with a non-deterministic framework when attempting to computationally model actual pressing problems in order to cope with the uncertainty existing within them.

3 Bayesian Networks

3.1 Probabilistic Modeling

As one non-deterministic approach, there are probabilistic methods. By using probability, it becomes possible to quantitatively model the non-determinism of phenomena and to treat it strictly by means of axiomatic probability theory. The probability values to be assigned to observable phenomena can be obtained from a large quantity of observational data; and for unobservable phenomena, estimates may be made by Bayesian probability theory (Bayesian hypothesis). This models the uncertainty of variables and the relationships between them through conditional probability and is easy to understand when considered as a way of determining the uncertainty of a particular variable, given information about other variables. Since this unknown probability distribution is treated as a subjective prior distribution in the conventional Bayesian hypothesis, it has been criticized by non-Bayesian statisticians. Recently, however, as large amounts of data have become more manageable, it has become possible to empirically construct this probability distribution from large statistical data sets and this approach is promising as a practical method in domains having a large amount of uncertainty.

For example, consider a stochastic framework for treating totally unobservable phenomena. There is a large amount of uncertain information in the real world, such as future weather conditions, noise signals, or user intentions, for which it is difficult to determine a specified value. We introduce a stochastic framework in order to systematically cope with these. An object that includes indefiniteness, such as complex factors or the influence of noise, will be represented as a random variable denoted by X, and the concrete values this variable can take will be represented by $x_1, x_2, \ldots, x_n$.

Next, consider dependencies between variables. For example, when it has been established that the variable X_j becomes y if X_i takes the value x, X_j can be considered to be dependent on X_i (if X_i=x then X_j=y). When considering complex phenomena that actually occur, the dependencies among multiple variables become complicated, and explicitly enumerating all relationships, such as “if $X_1=x_1,\ldots, X_i=x_i, \ldots$, then X_j=y”, is not very realistic. In addition, even if this type of If-Then rule were enumerated extensively, in practice, there are exceptions, and always describing the situation completely is probably difficult. Therefore, we abandon the exact expression, focusing only on the primary variables; and in order for a rule to quantitatively demonstrate the extent of confidence, we introduce the following probabilistic expression: “when $X_i=x_i$, the probability that X_j=y is P (X_j=y | $X_i=x_i$).” A unique dependence between the two quantities x and y can be represented, for example, by the function y=f(x); and in the same way, the dependence between the random variables X_i and X_j can be represented by the conditional probability distribution $P(X_j|X_i)$.

This shows that the distribution for X_j is influenced by the values taken by X_i and that the quantitative version of this dependence is established by the conditional probability distribution $P(X_j|X_i)$. Further, quantitative dependence among multiple random variables can be modeled by a set of a graph structure and conditional probability tables defined on each variable, that is Bayesian network.

The fact that arbitrary variable probability distributions can be calculated efficiently, with no distinction between predictor variables and criterion variables, is also a strong point of the Bayesian network construct; and models can be reused in various applications.

A framework that determines model or system behavior by providing data consisting of groups of desired inputs and outputs is referred to as machine learning or statistical study. A Bayesian network can also be constructed through statistical studies from actual data. Calculations of probability distributions performed on Bayesian networks are called probabilistic inferences. Below is a simple discussion of the construction of models and probabilistic inferences from

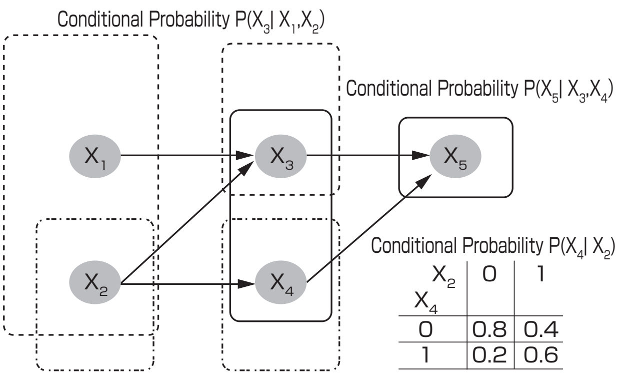

Fig. 1 Bayesian network.

models and data.

3.2 Bayesian Network Model

Mathematically, in Bayesian networks, a model is defined by a graph structure, which considers random variables to be nodes, and in which a conditional probability distribution is allotted to each node (Fig. 1).

In the case of discrete random variables, the conditional probability distribution of each variable is given by means of a conditional probability table (CPT). Giving a table of conditional probabilities in this way allows the probability distribution to be expressed with more degrees of freedom than is the case by specifying a density function and a parameter. In other words, it is useful as a non-deterministic modeling procedure when the nature of the object is not known in advance.

Destination variables that give the conditional probability are referred to as child nodes. In this way, directed acyclic graphs defined by a conditional probability table, variables, and graph structures are constructed as Bayesian network models.

3.3 Model Construction from Model Data

When Bayesian network models become large, it is not easy to determine the network structure or the entire conditional probability table manual. In such cases, a procedure is necessary for constructing a model from statistical studies of large amounts of data.

Utilized data sets that include cases which deal with all items in the conditional probability table are called complete data. In this case, the statistical data is counted to obtain the frequencies; and these, when normalized, become the most likely estimators of the conditional probability values. In the case of incomplete data having deficiencies, conditional probability values are presumed, compensating for various types. There are instances when it is desirable to construct the model network from data. Studies of the construction then search for the graph structure from some initial conditions. As a measurement criterion for the appropriateness of a graph structure, information criteria other than likelihood, such as AIC, BIC, or MDL, etc., are used. When the graph node number is large, the search space increases explosively, and from a computational load perspective, searching all graph structures is difficult; therefore, it is necessary to use a greedy algorithm or various types of heuristics to search for quasi-optimal structures. The K-2 algorithm [5] is a study algorithm for this type of graph structure. This search algorithm is as follows:

Table 1 Conditional Probability Table (CPT).

$P(y_1\|Pa(X_j)=x_1)$	···	$P(y_1\|Pa(X_j)=x_m)$
:	···	:
$P(y_n\|Pa(X_j)=x_1)$	···	$P(y_n\|Pa(X_j)=x_m)$

(i) for each node, limit the candidates that can become new nodes, (ii) select a child node, add and graph the new candidate nodes one by one, (iii) decide on and evaluate the parameters on which the graph is based, (iv) only when evaluated highly, use as a new node, (v) when there are no more new node candidates to add, or when the evaluation does not increase even if a new candidate is added, move to another child node, (vi) repeat (i) – (v) for all child nodes.

In general, new search spaces increase combinatorically; therefore, a device is needed to avoid an increase in computational load by limiting combinations of new nodes that become ranked from the beginning to be candidates. Furthermore, we consider independently the search portion of the graph (ii), (v), and the mode evaluation portion (iii) and think about various study methods.

One can expect that the use of a Bayesian network would be an effective approach to construct a non-deterministic model from large amounts of data by means of statistical learning. However, obtaining a causal structure from only statistical data is fundamentally difficult, and the task of searching the graph structure is NP hard. In such cases, it is actually necessary to skillfully implement the variable candidates or search range limitation, or to introduce appropriate latent variables.

3.4 Probabilistic Inference

There are other types of models that possess a graph structure; however, many of them are often used to visualize graph structures that explain data. On the other hand, by constructing them with discrete random variables and conditional probability tables, Bayesian networks very efficiently implement the probabilistic inference algorithm, which estimates the probability distribution of arbitrary random variables in the model. This is a significant advantage over other graphical models and is a crucial feature in operating an intelligent learning system with a realistic computational load.

Probabilistic inference on Bayesian networks is implemented by the following procedure: i) assign the value of an observed variable (e) to a node, ii) assign a prior probability distribution to both new nodes and nodes having no observed value, iii) calculate the posterior probability distribution P(X|e) of the desired object value (X).

In order to find the posterior probability in item ii), a probability propagation method is implemented which renews the probability distribution of each variable according to the dependence between variables.

When all paths within a graph structure that does not consider the direction of links of the Bayesian network do not possess loops, the Bayesian network is called a *singly connected* network. In this case, even networks with structures in which multiple new nodes and child nodes exist, the calculation is completed by utilizing its conditionally independent character by performing, for each node, probability propagation calculations of 4 types: propagation upstream, downstream, from upstream, and from downstream (Figure 2).

These computations are completed in order of network size (number of links), and the calculation efficiency is extremely high. When the network is viewed without considering the direction of links and there is a portion in which even one path is somehow looping, this Bayesian network is called *multiply connected*. In this case, there is no guarantee of an exact solution; however, probability propagation can be applied as an approximate solution method; this is called the Loopy BP method.

4 User Modeling

Based on the fact that the information system and the user advance processing conversationally, and that this information system is a portion of the entire system, which is subject to operations, the information system, the user, and even the environment and surrounding circumstances, must be considered. Consequently, viewing the entire system as a control object, human behavior and response should be thought of as one component of the object of calculation. In this case, we want to evaluate what the user is requesting under given circumstances and how to react to the system output results obtained. In the system, it is necessary to describe the cognitive state of such users in terms of computable user models.

Machines (programs) learn from data by the development of machine learning. In other words, the approach wherein models are constructed from data and revisions are made sequentially [iteratively] is feasible. In the construction of machine learning models, statistical tests are repeatedly implemented to act as an automated model selection process, based on information criteria; and the result is considered to be the appropriate model. In other words, a statistically meaningful model is chosen through machine learning from within an extensive search space.

$$\Pr(X=x)=\alpha\lambda(x)\pi(x).$$
$$\pi(x)=\sum_u P(X\mid U=u)\prod_{Ui}\pi_{UiX}(Ui),$$
$$\lambda(x)=\prod_{Yj}\lambda_{YjX}(x),$$
$$\pi_{XYj}(x)=\pi(x)\prod_{k\neq j}\lambda_{YkX}(x),$$
$$\lambda_{XUi}(u)=\sum_x\lambda(x)\sum_{k\neq i}P(x\mid U)\prod_{k\neq i}\pi_{UkX}(u_k).$$

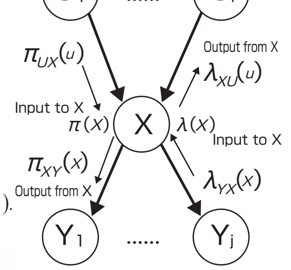

Fig. 2 Belief propagation.

Further, by using a probabilistic model, it is also possible to model abstract and diverse elements, including, for example, the idiosyncrasies of individuals, making it possible for conventional psychology to attempt to deal with a universal human model. This is an important viewpoint for models that implement individual adaptation or personalization crucial in information processing associated with recent human centered design and usability aspects. There have been various implementations of user modeling which use a Bayesian network as the probabilistic model [6]. In order to model, in particular, constructions of human cognition and evaluation as Bayesian networks, interview methods used in clinical psychology or marketing are applied [7]. In this way it is possible to infer acceptability or intention by modeling and implementing probabilistic inferences of the system and services of the user.

5 Modeling of Everyday Activities

As an actual example of computing uncertainty, the standpoint of user modeling has been discussed above; however, when considering everyday life assistance as various actual services[8], modeling of the living person, each day, as a user, is vital. Until now, sensors have been installed in a home, or a “sensor house” has been proposed for research and development in order to analyze everyday activities [9]. Until now, several applications have been proposed wherein abnormalities are judged by detecting outlying values while modeling patterns in measured data as stationary distributions; however, for broader applications, modeling with only stationary distributions is inadequate. Rather, it is necessary to consider optimizing the utility, or value, depending upon user intentions. In other words, in order to predict the intentions, impressions, and assessments of the user, which cannot be observed directly, from observable actions, higher-order inference is necessary. To accomplish this it is necessary to model dependencies and causal relationships, such as how results turn out in response to certain circumstances and activities; to accomplish this, it is necessary to record, not only action data, but comprehensive data involving causal variables, and to search for causal structure from relationships among a large number of variables.

This can be considered as a new kind of analysis of behavior, opened up by sensor technology and modeling technology. Behavior analysis was established by Skinner in the mid 1900s as a field of research making use of the behavioral science approach within psychology [10]. In this context, human activities are referred to as antecedent and behavior contingencies, and expected changes in environment, as the result of actions, are thought to be determined from relationships among three items. Further, the causal relationships between antecedents and behavior contingencies are clarified when focus is on a certain activity. While modeling this explicitly and by causing changes in those behavior contingencies and antecedents, control of behavior is implemented.

It is necessary to interpret video images of observed activities and to perform a labeling procedure in order to discover the cause and effect of activities. Performing this manually, however, requires an enormous amount of time and effort; therefore, it is difficult to efficiently analyze natural activities in an everyday life environment. In addition, performing this interpretation manually allows only a small number of objects to be analyzed as control variables of the activities. Technology capable of automatically handling large amounts of observational data is necessary in order to analyze everyday activities.

In such cases, observing actions automatically by a sensor network embedded in the environment, and utilizing statistical study techniques comes to mind. By constructing a Bayesian network model through statistical learning from the large amount of sensor data gathered, it is possible to connect the inevitability amid the reasons and purposes for actions, which become candidate behavior contingencies, and the environment and situations, which become antecedents. In this way, it is expected that behavior analysis will be developed largely by contributions from model construction technology that extracts variables with strong causal relationships from sensor technology that can comprehensively observe daily activities, and from the data observed thereby. Through modeling of actions based on Bayesian networks and ultrasonic sensor networks, research has, until now, been aimed at analysis of everyday life behavior [11] and at applications such as injury prevention for children [12][13]. An example of inferring the behavior of children [14] is introduced below.

Attach an ultrasonic transmitter to a person or object in the room. Then, at regular intervals, position information (x, y, z coordinate data) of the person or the objects can be captured by ultrasonic receivers embedded in the sensor room. A fisheye camera (camera with a wide-angle lens) installed on the ceiling of the room simultaneously photographs the circumstances of the person’s activity in the room as a video. For the activities of the person in this photographed room, the video images are labeled manually at one-second intervals. For example, a detailed database is collected giving action labels such as “the person is walking,” “sitting,” “standing.” Modeling of everyday activities is performed using this data, and experiments of behavior inference are performed based on the video.

Considering the problem as the system’s observation of the behavior by means of sensors and images, it can be formulated as a type of pattern recognition problem. Since data arising in the everyday real world considers human life activities and the living environment as a background, the nature of the state space in which the data arises and the bias are reflected in ways that resonate with human activity and semantics. In the space in which this type of data arises, peculiar restrictions and deviations in the frequencies of occurrence can be treated in terms of a probability distribution. Enumerating all the causal structures established in the world, as is the case for physical laws, is difficult because of the quantity of such descriptions; however, expressing the important elements as probabilities is an effective approximation. Work has been done to model this type of probabilistic construction by a Bayesian network in the actual space and to utilize it in Bayesian inference [15].

In Bayesian inference, multiple class labels are taken to be C_i, and a posterior distribution combining both the likelihood $P(x|C_i)$ for a signal pattern x and prior distribution $P(C_i)$ which determines the class label C_i such that

$$P(C_i|x) = P(x|C_i)\, P(C_i) \,/\, \sum_j P(x|C_j)\, P(C_j) \qquad (1)$$

is maximized. It is known that this makes possible optimal recognition that minimizes the Bayes error probability. The fit of the data is represented by the likelihood, and prior knowledge is represented by the prior probability distribution. Learning from data and prior knowledge are naturally integrated by considering the maximized prior probability, which is the product of both of these, to be the inferred result.

In cases when the frequency of occurrence of class labels depends upon the observation time and place, the prior distribution $P(C_i)$ depends on the situation S. In such cases, consider this to be the conditional probability $P(C_i|S)$, replace this with $P(C_i)$ of equation (1), and obtain a class which maximizes the posterior probability of (2):

$$P(C_i|x, S) = P(x|C_i)\, P(C_i|S) \,/\, \Sigma_j P(x|C_j)\, P(C_j|S) \qquad (2)$$

The second term in the denominator of the right-hand side of Eq. (2) is the prior probability of the activity label C_i, which is in situation S in the label space. Here, we will consider the stochastic causal structure in the label space. When we construct, as a Bayesian network, a causal structure among a series of places and actions, for example, when we introduce advance knowledge with a causal structure of the following form: “If activity C_i^t occurred in circumstance S at time t, it is easy for activity C_i^{t+1} to occur at time t+1,” the probability of an action when a person enters domain S is expressed as $P(C_i^{t+1} \mid C_S^t, S)$ and can be modeled by a Bayesian network. Having constructed the model by means of statistical studies on the data set of observations of activities when children are playing in an experimental environment imitating a living room, aside from past activities, dependencies between the relative distance of, for example, the sofa or wall in the room, the speed of movement, etc. are confirmed. Having studied the Bayesian network and naïve Bayes by means of activity data of other children, and inferred the activities of other children through Bayesian inference via Eq. (2), the identification rate was found to be approximately 50 % or less according to the most likely inference of naïve Bayes only, and could be increased to approximately 60 %~80 % by Bayesian inference, using a Bayesian network [14]. By means of this behavior inference algorithm, it is possible to efficiently form action-labeled data from observational images of everyday activities.

6 Research as a Service

Now that large-scale data can be measured in daily life, complicated problems can be handled through statistical learning. However, a characteristic problem of statistical learning is that as models become complicated at a high level, the amount of data necessary for learning increases. Sensor data observable superficially can be dealt with comparatively easily. However, the internal state of human behavior is a psychological aspect; therefore, a questionnaire survey used on test subjects is a necessity, and this entails a high cost. In addition, when acquiring data, practical problems exist, such as the problem of privacy and the fact that cooperation for the purpose of the research is simply difficult to obtain. Furthermore, even if a phenomenon is easily observable in terms of external factors, in order to completely collect predictor variables with high environmental dependence at the scene where they will actually be used, it is necessary that the environment wherein data is observed be controlled, so as to simulate the everyday environment as accurately as possible. Therefore, for this type of problem, the author considers it obligatory to unify actual service, investigation, and research. In this connection, the author lectures on the concept of “Research as a service” [22]. This clarifies the “means-end chain” as behavioral contingencies of humans in the context of behavior analysis, making it easy to make comprehensive models while including environmental dependence. Consequently, the results of the observations, evaluation questionnaires, and user feedback (psychological investigation), obtained while implementing the information service in society, are collected without separating the investigation and modeling procedure from the applications that use the model. This is known historically in cybernetics and in reliability engineering as the Deming cycle: PDCA (Plan, Do, Check, Action), in which a model is continuously corrected while cycling through actual problems.

For an essential resolution of the uncertainty issue, an approach is necessary in which a cycle is permanently continued that collects additional data while using and controlling the model, with modeling based primarily on actual data. This is not limited to simply collecting actual data, but from the standpoint of research, implies that the researcher is imbedded in the field, which leads the way to new research that will bring about new values and evaluations [16].

It is also necessary to implement an applied system that can be embedded as infrastructure within society, as a product that can be tolerated as an actual service.

7 Bayesian Network Applied System

The applied system of the Bayesian network can be developed by implementing a probabilistic inference algorithm or model construction algorithm as a computer program. Building on work in the Real World Computing Project, the IPA Unexplored Software Project, and other such projects prior to 2001, and by searching Bayesian networks from large amounts of data in 2002, the author developed the software BayoNet that can perform probabilistic inference based on this work [17][18]. This software has been licensed to private enterprise and commercialized; however, due to the fact that a high degree of specialized knowledge is necessary in order to apply it to the resolution of particular problems and the fact that the utilization procedure is not self-explanatory, it has been somewhat difficult to train users who can fully utilize the software. If it is software developed for highly specified purposes, it is not necessary; however, software featuring a Bayesian network emerging as purely fundamental research on mathematical models can be applied for extremely broad purposes; and at the point in time when it can be utilized in practice, new investigations resulting in even more valuable purposes can occur. Therefore, a taskforce of the strategic center for venture development was started in 2003, and researchers personally had the opportunity to begin a search for business models that use this technology. At this point in time, many research results have remained as essential technologies, such as algorithm refinement and acceleration or inference precision improvements; however, we have felt resistance to further development of technologies in circumstances in which the outcomes were not obvious. Therefore, we decided to prioritize the search for outcomes by problem resolutions that had the possibility of being adequately treated, given the efficiency at the time.

The advantage of using Bayesian networks is that by performing probabilistic inferences, we can determine the probability distribution of arbitrary variables and conduct quantitative evaluations in various situations. In many conventional multivariate analysis procedures, quantitative relationships are often modeled based on a covariance relationship that assumes linearity among variables (linear independence). In the Bayesian network model, quantitative relationships are represented by a conditional probability table. In a conditional probability table, a family of conditional probability distributions are not hypothesized, but rather, the table forms a model in which non-linear, non-normal relation interactions can be represented with great freedom. In addition, predictor variables and objective variables are not clearly distinguished; therefore, introduction of latent [implicit] variables is also straightforward. In other words, even variables that cannot be observed can be treated as latent variables. Therefore, latent variables are introduced that become categories, and when analyzing the statistical data of a user or customer, we can extract attributes of groups that perform the same activities, classify constituencies, and that can even be utilized in customer segmentation.

It is extremely important that these characteristics respond by recommending information or products that are acceptable matches, depending upon the user or customer activities (Web browsing history, etc.), attributes, or circumstances. In collaborative filtering, information or products desired by customers or users cannot reflect situation dependence when displayed by a portable telephone or car navigation system. Information recommendation technology for such activities that change depending upon the environment is important, even in ubiquitous computing, in which a variety of situation changes in actual space are imagined.

7.1 User- and Situation-Dependent Information Recommendation in a Car Navigation System

It sometimes happens that the driver of a car wants to stop somewhere while driving. For example, while driving for some purpose, the driver decides to stop to eat at a restaurant. In conventional car navigation systems, a category is specified, and all corresponding restaurants are listed in order by distance. The user must find the appropriate restaurant from within the list; however, the user has to operate a touch switch or remote control in order to see detailed information about restaurants, so it is not easy for the driver to locate the desired restaurant.

Therefore, if a car navigation system were to model the driver's preference of various restaurants, given various situations and criteria using a Bayesian network, and, using a probabilistic inference from this, if the system replaced the driver while driving, and automatically selected the appropriate destination, it would be an extremely practical function. A person's taste depends largely upon their personality and upon the situation while driving. While driving, it is necessary to select the most appropriate choice at the time, among conditions that change moment by moment.

To illustrate this dependence on situations and personal differences, a Bayesian network can be efficiently applied that can model complex relationships among variables and uncertainty. Therefore, we test and evaluate a car navigation system that suggests content appropriate for the user [6]. This system possesses, as a Bayesian network, a user taste model within the vehicle information system. Content, such as restaurants or music, is suggested by content providers, and a score showing how appropriate it is for the user and conditions at the time is calculated as a conditional probability when the situation and user attributes are given.

It then recommends items with a high score, limiting them to superior content. For 182 actual restaurants in the Shinagawa neighborhood, a questionnaire was conducted among 300 test subjects, causing desired store locations to be selected in six situations (scenarios). A model was constructed from the gathered data. Restaurants desired in six situations (scenarios) were selected from the 182 restaurants in the Shinagawa neighborhood. Concerning the selection procedure, firstly, the user was queried about desired categories, and stores corresponding to those categories were displayed. If disliked, the next genre was chosen by the same selection method as in currently existing car navigation systems. There were multiple answers for selected restaurants, and ultimately 3778 records were obtained. There were 12 situation attributes, 17 restaurant attributes, and 12 user attributes. The model in Fig. 3 was constructed as a result. There are four attribute nodes representing users, three representing situations, and six representing restaurants. The model consists of all 13 of these random variables, and the probability distribution of restaurant attributes favored by specific users in a given situation is calculated by probabilistic inference.

In the model of Fig. 3, for drivers with a light driving history, the probability is high that franchise restaurants such as family restaurants and fast food chains will be chosen; conversely, for extensive driving histories, the probability that these restaurants will be chosen is low. Franchise restaurants often provide parking areas and show a tendency to be favored by young or beginning drivers. In addition to "driving history", there is a "have plans" interaction. This reflects the tendency that even in cases wherein the driving history may be long, in situations when the driver has plans and must hurry, there is a high probability that a franchise restaurant will be used. The proper tendency is obtained intuitively for other relationships, such as that between budget level and vehicle type.

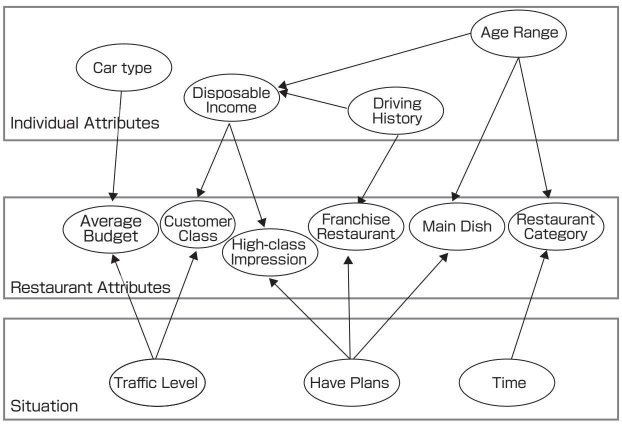

Fig. 3 Restaurant preference Bayesian network model.

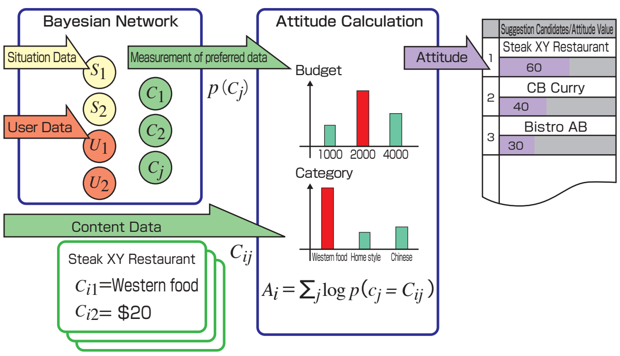

Fig. 4 Outline of restaurant recommendation system [6].

Using the model depicted in Fig. 3, a prototype of a restaurant recommendation system was also designed (Fig. 4). Favored content attributes are forecast as probability distributions from user variables and situation variables.

$$A_i = \sum_{j=1}^{n} {}_j \log p\,(\,c_j = C_{ij}\,) \tag{3}$$

By recommending content for which the value of this score is high, a car navigation system appropriate for the situation and the user can be implemented. Upon comparing this prototype system and a conventional car navigation system, its effectiveness was confirmed by the fact that prediction results for restaurants matched the users' preferences and the situation.

7.2 Information Recommendation Appropriate for User and Situation with a Portable Phone

Information recommendation technology appropriate for various users and situations is important in next generation portable phone services. Examples of application of Bayesian networks in a movie recommendation service in portable phone services have been introduced [19][20]. For approximately 1600 test subjects, their content evaluation history, user and content attributes were collected via a questionnaire that suggested movie content. Other than demographic attributes such as age, gender, employment, etc., questions regarding lifestyle, appreciation frequency as attitude attributes concerning movie viewing, concern over movie selection time, the primary purpose for watching movies (seven questions about wanting to be emotionally moved), evaluation of content (good/bad), one's mood at the

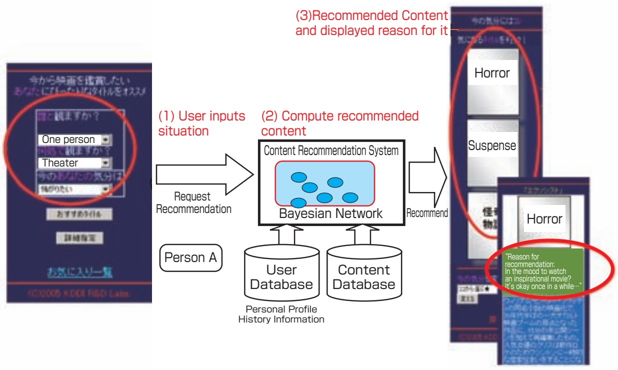

Fig. 5 Mobile information service system that recommends movies depending upon the user and situation [19][20].

time (seven questions about being emotionally moved) were collected. Furthermore, for approximately 1000 people, all of the following were collected separately as free-form text: the content of each movie, what kind of feeling or situation, (theater, DVD, etc.), with who, with how many people, what time of day, was the movie appreciated. This data was input into BayoNet [17][18], the Bayesian network construction software developed by the author, and a Bayesian network model was constructed automatically. Through the Bayesian network constructed in this way, a prototype of the portable information system was developed that makes movie recommendations, based on situation and user tastes. If the user sends requests to services from the portable phone, together with information about the situation, the system implements the probability inference using registered user attribute and situation information from the database. Content whose probability of being selected is judged to be high is recommended as superior (Fig. 5). This movie recommendation system was also developed into an Internet service at *auOne* lab (http://labs.auone.jp) and released generally in 2007 with approximately 7000 recommendations implemented. Further, the model is being restudied from this recommendation history, and experiments are being conducted to improve recommendation precision. Using the calculation model for movie selection constructed in this way, we also proposed cooperation with a movie distribution company to optimize sales strategies for DVD content for which the movie release period has passed [21].

As this information service spreads and multiple users utilize the system, the history of selection content accumulates ever-larger amounts of statistical data. Improvements in the Bayesian network model resulting from that data will increase the appropriateness and inference precision of the model, create a self-supporting feedback loop, and allow horizontal development of other services to be realized. Data obtained from the market through actual services becomes reusable knowledge for the calculation model; this knowledge cycle, reflected in the next service, can be called "Research as a Service" as noted earlier (Fig. 6). Research activities through this type of actual service can even be put into practice in a service engineering research center through construction of a calculation model from large-scale data and through implementation of an optimization design loop in the field. Such research activities are proposed as a business to improve the productivity of the service industry [22].

8 Conclusion

In the present research, the development of software could be categorized as pure or basic research; however, software development which excludes the initial step, which is obviously outcome oriented, could be considered applied fundamental research. It seems that there were several conditions that implied that we rethink the criterion for application selection intuitively recommended in that process.

1. There are unresolved problems in existing procedures.
2. There were problems actualized by user requests.
3. There are stakeholders that profit from resolutions of these problems and bear the corresponding cost and risk.

In these types of conditions, using Bayesian networks to model human behavior, forecasting customer or user activities, and achieving improvements in value and efficiency by optimization of associated services is thought to be an appropriate outcome. Client enterprises that can realize these outcomes exist in industry types that possess contact points (channels) with various customers. Selecting the outcomes mentioned above, the appropriate fields become channels that can collect large amounts of data from customers such as the Internet, portable telephones, car navigation systems, and call centers. However, among these choices, two necessary types are: being able to

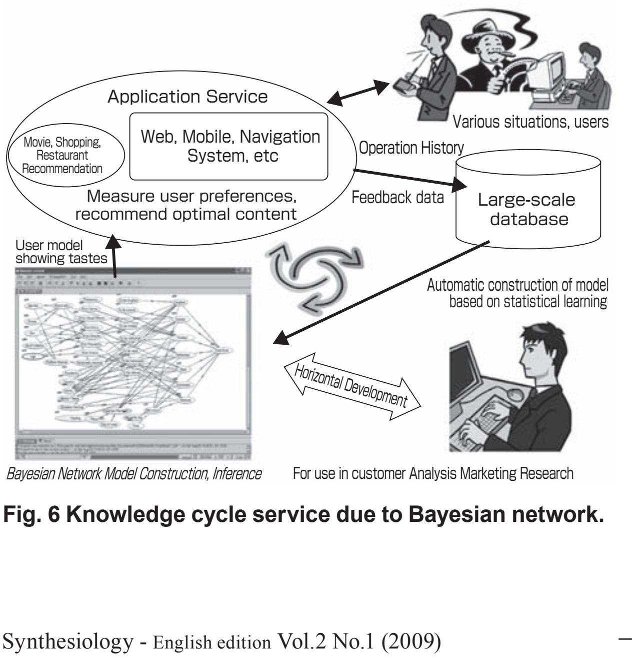

Fig. 6 Knowledge cycle service due to Bayesian network.

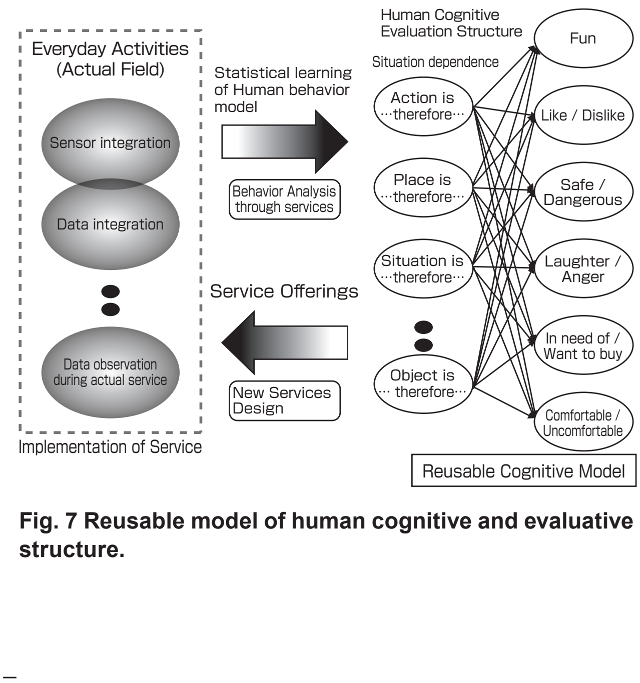

Fig. 7 Reusable model of human cognitive and evaluative structure.

adequately respond by transferring the present technology, and the development of additional technology for outcome realization. In the former, the venture responds; in the latter, the choice is made to promote cooperative research between AIST and enterprises.

Engineering implementation and societal implementation differ. In engineering, even if the technology is already established, in order to produce societal value, participation of many more stakeholders is necessary. It is necessary to convey value to these stakeholders, which will not necessarily have an engineering background, in order to persuade them to bear the cost and risk; and it is necessary to demonstrate that the outcome has high reliability. Therefore, societal implementation through department-level cooperative research and technology transfer to AIST Ventures is necessary, and results need to be proven in the field. In other words, evaluation of the outcome and societal implementation occurs simultaneously.

In order to clarify the conditions under which implementation in society is possible, a marketing research was performed in the Venture Task Force. The cost benefit analysis, which did not need consideration in the first type of basic research, was critical. In order to smoothly advance societal implementation, reductions of cost and risk are sought while improving benefits. At this step, the outcome itself is corrected, and there is a possibility of motivating fundamental research out of necessity for a new outcome. This promotes fundamental research, becomes feedback to fledgling basic research, and is represented in the policy statement of the Digital Human Research Center: “Application driven fundamental research.” It has also become possible to acquire large-scale data that includes situations and context involving the results of activities through actual services and actual users. Bayesian networks constructed from this data forecast the cognitive and evaluation structures and behaviors of existing consumers and others. Being causal models rather than merely descriptive models of the data, they are cognitive models with high reusability and potential for horizontal development (Fig. 7) in other services [23].

Concerning issues required for implementation in society, and from the standpoint of fundamental research, whether or not a quick response can be given is thought to be an important issue associated with establishing fundamental research on problem resolutions requested by society in the future. The very fact that speed is requested of technology in society requires that many “buds” be nurtured for the future. Such choices can only be performed by those thoroughly grounded in fundamental research; consequently, in order to perform fundamental research, views aimed at societal technology that clearly envisions the future are surely required.

References

[1] S. Russell and P. Norvig: Artificial Intelligence, A Modern Approach, Prentice Hall Series (2002).

[2] J. Pearl: *Probabilistic inference and expert systems*, Morgan Kaufmann,CA, (1988).

[3] P. Baldi, P. Frasconi and P.Smyth: *Modeling the internet and the web – probabilistic methods and algorithms*, (2003).

[4] D. Marr: *Vision: A computational investigation into the human representation and processing of visual information*, W.H.Freeman and Company (1982).

[5] G. Cooper and E. Herskovits: A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*. 9(4), 309-347 (1992).

[6] Y. Motomura, T. Iwasaki: *Beijian Nettowaku Tekunoroji*, Tokyo Denki University. press (2006) (in Japanese).

[7] Y. Motomura and T. Kanade: Probabilistic human modeling based on personal construct theory, *Journal of Robotics and Mechatronics*, 17 (6), 689-696 (2005).

[8] Y.Motomura, Y.Nishida: Behavior understanding for everyday support systems in daily environment, *Journal of Japanese Society for Artificial Intelligence*, 20 (5), 587-594 (2005) (in Japanese).

[9] M.Minoh: Human daily life support at a ubiquitous computing home, *Journal of Japanese Society for Artificial Intelligence*, 20 (5), 579-586 (2005) (in Japanese).

[10] B. F. Skinner: *Behavior of Organisms*, Appleton-Century-Crofts (1938).

[11] K.Shiraishi, Y.Yasukawa, Y.Nishida, Y.Motomura, H.Mizoguchi: Information Management Systems for understanding everyday life activity, *Annual Conference on Japanese Society for Artificial Intelligence*, 3G3-03 (2008) (in Japanese).

[12] Y. Nishida, Y. Motomura, G. Kawakami, N. Matsumoto and H. Mizoguchi: Spatio-temporal semantic map for acquiring and retargeting knowledge on everyday life behavior, *Lecture Notes in Artificial Intelligence, JSAI 2007 Conference and Workshops, Revised Selected papers*, 63-75, Springer-Verlag (2008).

[13] G.Kawakami, Y.Nishida, Y.Motomura, H.Mizugochi: Behavior modeling base on spatio-temporal expansion of behavior metrics sensing, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, 20 (2), 190-200 (2008) (in Japanese).

[14] S.Ishikawa, Y.Motomura, S.Kawata, Y.Nishida, K.Hara: Inference and construction of probabilistic causal structure model for everyday life behavior, *Annual Conference on Japanese Society for Artificial Intelligence*, 3G3-04 (2008) (in Japanese).

[15] Y.Motomura, Y.Nishida: Graphical modeling of prior probability in Bayesian estimation for behavior understanding in real life, *IPSJ Transactions on Computer Vision and Image Understanding*, 18, 43-56 (2007) (in Japanese).

[16] Japanese Science and Technology Agency, Center for Research and Development Strategy report, CRDS-FY2005-WR-16(2007) (in Japanese).

[17] Y. Motomura: BAYONET: Bayesian network on neural network, *Foundation of Real-World Intelligence*, 28-37, CSLI calfornia, (2001).

[18] Y.Motomura: Bayesian network software BayoNet, SICE Journal of Control, Measurement, and System Integration, 42 (8), 693-694 (2003) (in Japanese).

[19] C. Ono, M. Kurokawa, Y. Motomura and H. Asoh: A context-aware movie preference model using a Bayesian

network for recommendation and promotion, *Proc. of User Modeling 2007*, *LNCS*, 4511, 257-266, Springer, (2007).
[20] C.Ono, Y.Motomura, H.Asoh: Context-aware preference handling technologies on mobile devices, *Journal of Information Processing Society of Japan,* 48 (9), 989-994 (2007) (in Japanese).
[21] K.Ochiai, T.Shimokado, C.Ono, H.Asoh, Y.Motomura: Supporting movie contents marketing using Bayesian networks, *Annual Conference on Japanese Society for Artificial Intelligence* (2009) (in Japanese).
[22] Y.Motomura et.al.: Large scale data observation, modeling, service design and applying loop for service innovation, *Journal of Japanese Society for Artificial Intelligence*, 23 (6), 736-742 (2008) (in Japanse).
[23] Y.Motomura, Y.Nishida: Difficulty of behavior understanding research and a knowledge sharing framework, *Annual Conference on Japanese Society for Artificial Intelligence*, (2007) (in Japanese).

Author

Yoichi Motomura
The team leader of large scale data based modeling research team, center for service research(CfSR) and the senior research scientist of digital human research center (DHRC) at National Institute of Advanced Industrial Science and Technology (AIST) in Japan. He received M.Sc from The University of Electro-Communications and then joined the Real World Computing project in the electrotechnical laboratory, AIST, MITI (the ministry of international trade and industry, Japan) in 1993. In this project, he developed Bayesian network modeling and inference system, BayoNet. His current research works are statistical learning theory, probabilistic inference algorithms on Bayesian networks, and their applications to user modeling and human behavior understanding and prediction from sensory data. He also conducts the venture company, Modellize Inc.,as the chief technology officer and applies Bayesian network technologies to many variety applications in Japan. He also wrote four textbooks on Bayesian networks, and gave many lectures on Bayesian network researches and applications. He received a best presentation award, research promotive award from Japanese Society of Artificial Intelligence,Docomo Mobile Science award and IPA super creator award.

Discussion with Reviewers

1 Clear description of the originality as synthesiology

Question and comment (Hideyuki Nakashima)

The author described the difficulty of the modeling of daily-life behaviors and the significance of machine learning approaches for the modeling in chapters 1 and 2. The part should include more concrete examples for the improvement of readability of ordinary readers.

The author has described “Research as a service” in chapter 2, and I believe that this part is the most essential part of the presented paper. This part should be expanded.

Question and comment (Masaaki Mochimaru)

I guess the paper can be understood as presenting a breakthrough-type *Full Research*. The core technology of the paper is the Bayesian network with sensing technology and/or interview. The result of the study is a *Synthesiology* that realizes the solution of problems in the real world. The method and results are not just integration of the peripheral technology. The originality of the paper as *Synthesiology* is the suggestion of the concept of “Service as a research”. That is a social-circulation-type *Full Research*. This paper has embodied the concept of *Synthesiology* more than the past articles in *Synthesiology*, so, the author should explain the concept of *Synthesiology* explicitly.

Answer (Yoichi Motomura)

I have revised the abstract, title, the concept of “Service as a research”.

2 The composition of the paper

Question and comment (Masaaki Mochimaru)

This journal is not for artificial intelligence but *Synthesiology*, so, the author should write “the dream” realized through the integration technology in the introduction. It would be a description that the dream (the realization of daily-life-support-service by a system that understands the purpose of human behavior) and the concrete examples in order to give example images to readers. The breakthrough point for realization of the dream is “description, understanding and realization of daily-life in computer”. The difficulties of the realization of the dream are as follows: (1) human behaviors include some unclear and uncertain elements and the author has introduced the Bayesian network which is a non-deterministic modeling framework as a solution for the unclear and/or uncertain modeling, (2) the difficulty for using the Bayesian network is the necessity of a large scale dataset. This difficulty can be solved by the ubiquitous sensing technology and “Service as a research” in the real world. This composition can improve the readability of the paper.

Question and comment (Hideyuki Nakashima)

The logical connection between chapter 1 and chapter 2 is not good. Could you add a brief summary of the rest of the paper at the last part of chapter 1?

Question and comment (Masaaki Mochimaru)

The description of chapter 2 has prevented the smooth story expansion of the paper as indicated by Prof. Nakashima. I suggest three solutions as follows: (1) add the abstract to the last part of chapter 1 as say Prof, Nakashima, (2) change the position of chapter 2 and chapter 3, (3) move chapter 2 to the last of chapter 8.

I recommend the solution (3), because “selection of non-deterministic approach for human behavior model”, “Bayesian network as a non-deterministic approach” and “Research as a Service” are written in the paper. For readability it is not good to show the concept suddenly. So, I suggest the following storyline.

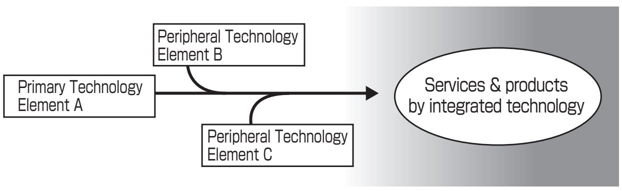

Fig. a Breakthrough model.

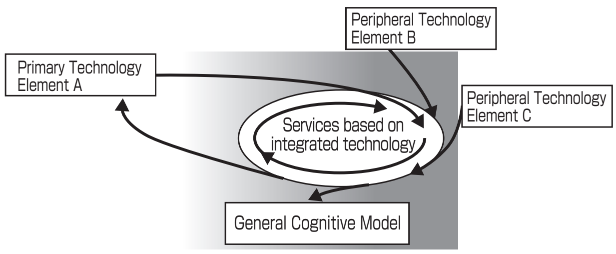

Fig. b Society cycle model.

(i) Description of "non-deterministic approach".
(ii) Description of the concrete examples.
(iii) Description of the need of a meaningful large scale data for realization of the examples.
(iv) Description of the proposal of "Research as a service" as the methodology of the study to obtain a large scale data.
(v) Description of the essentiality of cooperation of some stakeholders for the "Research as a service".

Answer (Yoichi Motomura)

Thank you for your suggestions. I have revised the composition based on proposal 1.

3 Additional description of specific examples

Question and comment (Hideyuki Nakashima)

"As this information service spreads and multiple users utilize the system, the history of selection content accumulates ever-larger amounts of statistical data. Improvements in the Bayesian network model resulting from that data will increase the appropriateness and inference precision of the model, create a self-supporting feedback loop, and allow horizontal development of other services to be realized", this part is the most important in the presented paper. Please describe more concretely about the spiral - what happened actually. I appreciate that you carried out the verified test of the proposed system to over 1000 people, however, you cannot say that the system is practical just by this. For example, it is usual to use the questioner dataset which is over 1000 samples in social science. In IT field, construction and evaluation of pre-production system is not enough to say that you actually used the system - To provide the actual service in the real world" is the significant step.

Answer (Yoichi Motomura)

I have added the description with respect to the examples.

4 Description of the reusable model

Question and comment (Masaaki Mochimaru)

The figure which shows the concrete image of "reusable model" corresponds to "general knowledge model" shown in comment 1. The interest as *Synthesiology* is in that the model can be horizontally developed into other applications (services). On the other hand, "the study is done through the service circulation in the real world. Additionally, the reusable model is generated". This concept's description maybe excess information for readers, and readability is not good. So, I suggest that the author describes the concept in chapter 8 with a figure.

Answer (Yoichi Motomura)

I have revised chapter 2 and chapter 8.