

CONCEPT SHARING BETWEEN HUMAN AND INTERFACE AGENT UNDER TIME CRITICALITY

Tetsuo Sawaragi¹ and Teruyuki Ogura²

¹ Dept. of Precision Engineering, Graduate School of Eng., Kyoto University

Email. sawaragi@prec.kyoto-u.ac.jp

² Sumitomo Electric Industries

Email. ogura@sei.co.jp

In this paper, instead of designing an automation system that is intended to replace a human operator completely with it and to exclude him/her out of the loop, we introduce an idea of an interface agent as a sophisticated associate for a human operator. Since its decision making style must be close to the human's proficient naturalistic decision making, we formalize its resource-bounded reasoning under critical time pressure by joining a machine learning method for concept induction and a classical decision theory.

1. INTRODUCTION

Recent popular concept of *human-centered automation design* has stressed the importance of the design philosophy of "people are in charge" or "human-in-the-loop" (Rouse, 1988). For this *idealistic* and very broad concept, however, Sheridan presented ten alternative meanings with a restrictive qualification saying that the real potential is somewhat questionable (Sheridan, 1997). As Sheridan depicted, we would like to stress that a human-centered design principle has to be able to answer to the following critical issues;

- how to make the human-autonomy and the mechanical autonomy (*i.e.*, automation systems) coexist letting them keep a friendly and sharable partnership, and
- how to avoid the human from the flood of data from the plant and of computer-based advice.

In the following of this paper, we at first introduce an idea of an interface agent as an alternative to the conventional stand-alone, isolated automation systems. Then, we will characterize that such an interface agent's reasoning tasks and its decision-making styles are quite different from the ones of conventional expert systems and from conventional decision support systems in that they are severely bounded to the *real-time contexts*. This is close to the human's *naturalistic* decision making style called *recognition-primed decision-making* (Klein *et al.*, 1993; Klein, 1993). We present a methodology for designing such an interface agent and formalize its

resource-bounded reasoning under critical time pressure by joining a machine learning method and a classical decision theory.

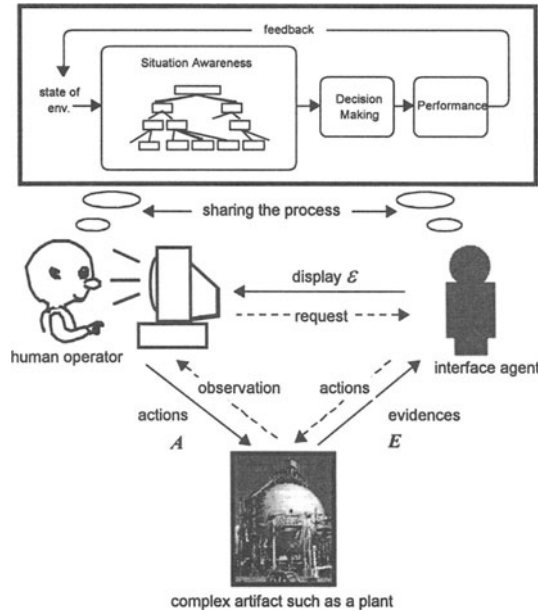


Figure 1 An interface agent as an associate for a human operator

2. INTERFACE AGENT AS A HUMAN ASSOCIATE

An interface agent is a semi-intelligent computer programs that can learn by continuously "looking over the shoulder" of the user as he/she is performing actions against some complex artifacts. It is expected to be capable of providing the users with adaptive aiding as well as of alternating the activities instead of a human (Maes, 1993; Maes, 1994). In this sense, an agent has to *coexist* with a human user so that it can evolve by itself as a human user's proficient level improves. It also has to be able to stimulate a human user's creativeness coordinately by changing its role dynamically as a human's associate, rather than to replace the human user with itself.

To let an interface agent be a *human associate*, the most fundamental issue is to guarantee *transparency* of its internal reasoning activity to the human user and to have continuous opportunities to *interact* with the human user to share the *process*, not the *product*, of reasoning with him/her. One of the major drawbacks of the current automation is that automation masks its internal activities from the human user and turns out to be a black box. This is causing intransparency to the human and lack of variability in ways of interacting with the automation.

In order to grade up an interface agent from an automated agent to a human collaborative agent, we formulate a decision-making activity of an interface agent analogous to a human. Within current decision research fields, a shift from a classical normative decision making paradigm toward a naturalistic decision-making

paradigm is emerging (Klein *et al.*, 1993). The latter has concentrated increasingly on the proficient experts' *situation assessment* ability and their ways of looking at a situation and quickly interpreting it using their highly organized base of relevant knowledge. That is, these are abilities to recognize and appropriately classify a situation. Hereafter, we call this style of decision-making as a *recognition-primed decision (RPD)* model after Klein (Klein, 1993). The distinguishing feature of the RPD model is to attempt to describe how people bring their experience to bear on a decision and how they are able to continually prepare to initiate workable, timely, and cost effective actions. Especially we are interested in modeling their capability to act proficiently under severe time pressure (*i.e.*, under emergency); to identify the situation quickly and accurately and to act promptly with less time and effort to act.

In the following of this paper, we at first propose a recognition-primed decision model of an interface agent, that is embedded *inside* the human-artifact interactions and that has to work as an intelligent associate for a human user/operator in a time-critical situation. This is schematically illustrated in Figure 1. We put an emphasis on the characteristics of an agent's capability of *dynamical recognition* of the complex artifact to be controlled or to be monitored. Wherein, "dynamics" is caused by situational factors such as the emergency and temporal availability of evidences. We show such a human-friendly way of reasoning is comprehensive enough for a human user to share the process with the agent in a collaborative fashion. Then, this formulation is applied to a design of an interface for a teleoperated mobile robot.

3. DYNAMIC CATEGORIZATION OF PLANT ANOMALIES

Important aspect of our interface agent is how to organize an "appearances" of the world (*i.e.*, a complex plant), which is different from a mere collection of objective features but is specific to a particular human operator depending upon current time-criticality he/she is forced to work. To be a human-friendly associate for a human operator, an interface agent has to be able to present the current status of the plant so that it can help his/her situation awareness and can remind him/her of the recovering operation to be adopted (Sawaragi *et al.*, 1996). In time-critical, high-stakes situations, the time required by people to review information, and confusion arising in attempts to process large amounts of data, can lead to costly delays and errors.

Under such a time-critical situation, an interface agent must be able to flexibly organize the appropriate appearance of the plant status discriminating among what is *now* relevant and what is *not* for assisting an operator's *situation awareness*. In our system, a taxonomy of all possible plant anomalies is organized in a hierarchical fashion using a machine learning technique called *concept formation* (Fisher, 1986) as shown in Figure 2 (Sawaragi *et al.*, 1996). Wherein, the root node represents a class of concepts covering all possible anomaly types, and the leaf nodes represent the individual anomalous cases (*i.e.*, a hierarchy via *is-a* and *subset-of* relations). In terms of this hierarchy, determination of the appearance of the plant concerns with how to determine the appropriate categorization of the plant anomalies out of the hierarchical taxonomy (Figure 2); to find a set of exclusive subtrees whose extensions are regarded as equivalent. The category of the hypothesis on the plausible anomaly types can be defined with a variety of abstraction levels within the hierarchical taxonomy. We call each possible categorization within the

taxonomy a *conceptual cover*, that is a categorization of all possible anomalies into mutually exclusive and exhaustive classes.

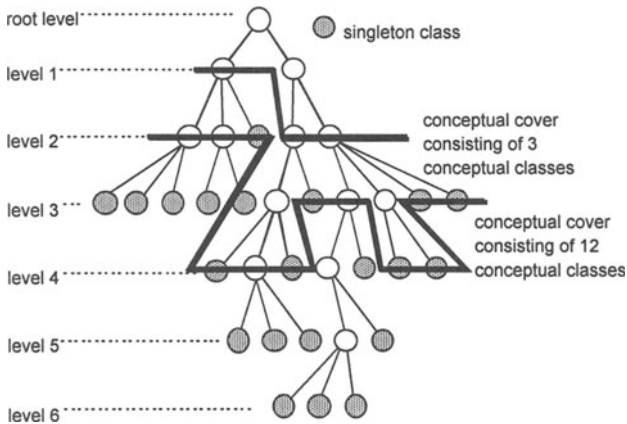


Figure 2 Conceptual cover of anomalous state concepts: a categorization adopted by an interface agent for its decision framing

4. RECOGNITION-PRIMED DECISION MAKING MODEL

In order to rationalize the agent's selecting the appropriate categorization to be presented to the human operator and also to be used for its own problem solving (*i.e.*, diagnosis), we at first consider about a general probabilistic reasoning model (*i.e.*, an *influence diagram* (Howard and Matheson, 1982)). We regard this as an agent's decision model that derives an appropriate action inferring the most plausible plant anomaly type by getting symptoms from evidential observations obtained so far. This is illustrated in Figure 3(a).

Wherein, a chance node H denotes a hypothesis node representing a state of the plant taking values of anomaly types. Nodes E^n 's represent a set of symptoms available to an interface agent, each of which corresponds to the features observable in individual instruments of the plant. As knowledge of plant anomalies, an agent has acquired dependency among the classes of anomaly types and their associating evidences as well as the prior probabilities of occurrence of those hypothetical classes. All of these are automatically acquired through the concept formation process, and have been stored in a taxonomy hierarchy (Sawaragi et al., 1994). Based on this *a priori* knowledge, getting new observations from the plant, an agent updates its prior belief on probable anomaly types based on a Bayes theorem. Then, these posterior beliefs are used for calculating the expected utilities of the available options defined in the decision node A based on the knowledge that are defined in the diagram defining the interrelationships among nodes of H , A , T and V in Figure 3(a). The option having the maximum expected utility would be determined as a recommendation for an interface agent to adopt. A diamond node is a special type of an oval node and is called a value node representing an agent's comprehensive utility. This shows that an agent's utility is determined by the types of anomaly and

the adopted action as well as by the delay. This reflects the fact that some anomaly types require immediate recovering operations as soon as possible and if it is delayed its utility would drastically decrease (*i.e.*, a *time-dependent* utility (Horvitz, 1991)). In our system, we categorize the utilities defined for all the combinations of hypothesis and actions into one of the three classes of time-dependent utilities according to the utility for the hard-deadline and the soft-deadline and the time-invariant utility.

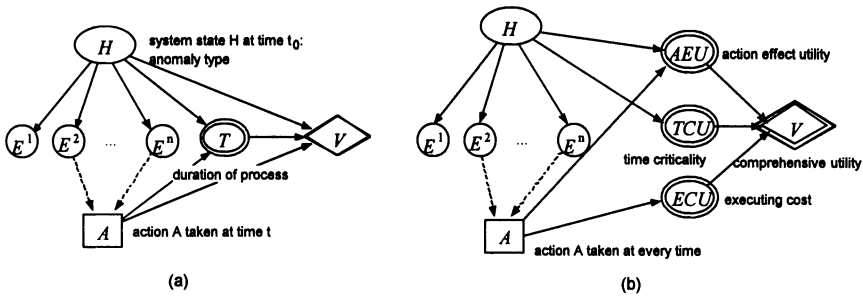


Figure 3 Interface agent's reasoning represented by an influence diagram

Figure 3(b) is another influence diagram used in our examples mentioned later, in which the value structure related to the value node V is more explicitly divided into the three subcategories of the decision maker's values.

In terms of these decision models, a definition of a domain of a hypothesis node H corresponds to a conceptual cover showing an appearance of the plant to an interface agent. This definition reflects the granularities of an agent's recognizing anomaly classifications (*i.e.*, a recognition-primed decision-making). The qualities of each possible conceptual cover must be evaluated in terms of the effectivity brought about by the adopted actions (*i.e.*, *expected value of categorization* (Poh *et al.*, 1996)).

Given a particular conceptual cover z , we denote the recommendation obtained by solving a model constructed using this by A^*_z , and define an expected value of categorization w.r.t. conceptual cover z as an expected utility calculated for the recommended option A^*_z and denote this by $EVC(z)$. Denoting a set of symptoms as a vector E and noticing the changes of utilities according the elapse of time t , A^*_z and $EVC(z)$ can be represented as follows;

$$A^*_z = \arg \text{Max}_A \sum_j u(A^*_z, H^f_j, t) p(H^f_j | E, \xi)$$

and

$$EVC(z) = \sum_j u(A^*_z, H^f_j, t) p(H^f_j | E, \xi),$$

where H^f_j represents a set of hypothesis classes of anomaly types making up the conceptual cover z , and $p(H^f_j | E, \xi)$ is the probability over them, given observations E and background state of information ξ . Note that for each conceptual covers z , a corresponding decision model may be constructed by retrieving the quantified conditional dependencies that are learned in the concept formation process and by joining them into a model. Thus, a conceptual cover z^* having the maximum EVC can be regarded as an action-oriented appearance of the world (*i.e.*, granular states) for an agent being affected by various situation factors at the current time t .

5. AGENT'S MANAGING COMPLEXITY UNDER TIME-CRITICALITY

As an intelligent associate, an agent aims to support an operator to focus his/her attention into the actual candidates of the anomalies rather than enforcing him/her execute a recommended action. That is, the agent presents the appropriate conceptual cover highlighting a set of possible anomalies while hiding its competitive other classes of anomalies behind (*i.e.*, distinction between "figures" and "grounds" in psychological terms). At the same time, this conceptual cover makes up the agent's decision model (*i.e.*, a hypothesis node H) that calculates the beliefs on the occurrence of possible. Here, we set the following criteria concerning with the determination of the appropriate conceptual cover:

- An agent should take into account of the expected delay of the operator's action execution in response to the presented information.
- An agent should aggregate anomalies into more abstract classes if the expected utilities of all possible actions for those anomalies are negligible.
- An agent should decompose a class of anomalies into more precise classifications if their probabilities of occurrence are high and the recommended action to be adopted varies depending on the anomalies within that class.
- An agent should avoid the risk of hiding the correct anomaly into the "grounds" and of highlighting the wrong one in the "figures".

Note that the above criteria are all reflected in the above-mentioned ways of determining the conceptual cover mediated by the agent's decision model.

The selection of the appropriate conceptual cover of this formulation was published elsewhere (Sawaragi and Katai, 1999).

6. NUMERICAL EXAMPLES

Based on the theory mentioned in the above, we present an example using well-known database in a machine learning field (*i.e.*, UC Irvine database). We use a database of animals consisting of 100 animal instances, each of which is represented by a set of 17 attributes with their binary values. As an analogy to the plant anomalies and other practical diagnosis, we newly defined three different actions (*i.e.*, approach, stop, and escape) available to the decision-maker (*i.e.*, an actor) that are commonly adopted against the individual instances of animals. That is, attributes correspond to the evidences available to the decision-maker, from which he/she has to identify what that animal is and to decide what action to take against that. With incomplete observation of those evidences, the decision-maker has to update his beliefs on the plausible hypothesis (*i.e.*, animal classes of the encountering animals). At that time, the role of the interface agent is to decide how to display information so that it can afford the decision-maker's easy recognition as well as his/her naturalistic responses properly. Note that the types of utilities shown in Figure 5 are assigned to each instance of animals with respect to each possible action alternatives, thus the difference of time-criticality is included in the model.

At first we construct an animal taxonomy using a conceptual formation technique of COBWEB as shown in Figure 4(c). Then, based on the information learned and stored within the individual classes of this taxonomy, we build an actor's decision model of Figure 3(b). Figure 4(a) shows the recommendations with their expected utility output by the decision models at each time a new evidence is added. This decision model is constructed by finding the most appropriate conceptual cover at each time according to the procedure mentioned in the previous section. Figure 4(b) represents a series of evidences that an agent obtains temporally. At each time an agent obtains evidences, it updates its beliefs on the hypotheses by Bayes reasoning being biased by the prior beliefs (*i.e.*, posterior beliefs in the preceding time step) and derives the appropriate conceptual cover at the current time. The shift of derived conceptual covers is shown in Figure 4(c), where the denoted filled box represents an instance revealing the evidences of Figure 4(b).

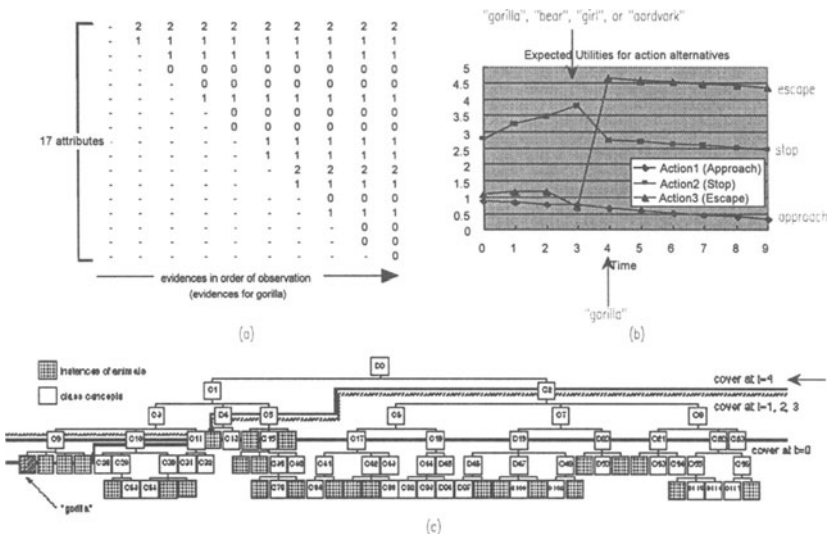


Figure 4 Shifts of confidence beliefs and conceptual covers according to the accumulation of evidences. (a) recommended actions (b) temporally observed evidences (c) conceptual covers

As shown in these figures, in the initial, the agent's recognition is still vague in a sense that it cannot focus into a particular animal. During the subsequent observations the agent's recognition is gradually confirmed, but the preciseness of the recognition is still coarse (*i.e.*, at the level of class C9). At this time the competitive hypotheses are gradually disconfirmed, and as a result, preciseness of their discrimination is not needed (*i.e.*, climbing up the hierarchy to the level consisting of C4, C5, C2). At this time the recommended action is quite *neutral*, *i.e.*, "stop" and its belief is still "fuzzy" meaning that the target may be either "gorilla", "bear", "girl", or "aardvark". But at time 4, a new evidence strongly confirming the animal is obtained, as a result, the agent's recognition turns out to be

more precise one (*i.e.*, going down the hierarchy from the level of *C9* to its descendants). Now that the target is strongly confirmed as the one of these siblings, the recommended action accordingly changes from the preceding "stop" to another alternative of "escape" as shown in Figure 4(a). Such alternation is due to the precise discrimination, or to the alternation of the "appearance" of the world of the agent, meaning that the target must be a "gorilla" and not a "bear", "girl", or "aardvark".

7. APPLICATION TO INTERFACE FOR A TELEOPERATED MOBILE ROBOT SYSTEM

7.1 Display Management for an Operator

We are now applying our proposing method to the interface for a teleoperated mobile robot system. As an automation device for a mobile robot, the so-called potential field method is well known. Fundamental issue for this approach is that the robot has to identify what the target object is; either obstacles to avoid or goals to approach. Obtaining evidences of the target object using the sensors on board, the robot has to identify that. These correspond to evidences and hypotheses in our formulation. Since the robot is moving with a constant velocity and has a limited computational resources for the sensing, this identification should be done within a finite time constraint, and the evidences can be obtained in a serial, temporal order, which is the same as discussed in the previous section. Different from the previous example, in this teleoperated mobile robot system, the robot may possibly encounter more than one target at the same time. For each target, the robot has to identify it respectively during a constrained time span. In this case, the robot attempts to identify each of them using respective conceptual covers within the common taxonomy (*i.e.*, a repertoire of the encountering objects).

We developed a display system that presents a robot's encountering environment to an human operator by showing a pseudo-3D virtual representation as illustrated in Figure 5. The cones distinguish the objects to be avoided or to be approached according to their colors, and the temporal beliefs as well as the identified abstraction levels within the current conceptual covers are shown by the differences of the heights and the transparency of the cones, respectively. Figure 5 illustrates how such a display changes when a robot obtains a new evidence. The conceptual covers shown to the right of the display illustrate the conceptual covers and the beliefs on the target object (*i.e.*, an object locating in immediate front) within that. In this way, an operator can perceive and monitor the change (*i.e.*, a process of an automated machine's reasoning) of a robot's semi-automated recognizing and moving tasks at the same time. This kind of concept sharing between the machine autonomy and the human autonomy will be essential towards establishing a balanced automation between a human and a machine.

7.2 Towards a Socially-Centered Automation

In my previous paper, we have proposed an idea of *Socially-Centered Automation* that regards an automated system as a true, equivalent partner of a human user

having an analogous capability of "sociality" to be embedded within the interactions with the human user. In contrast to a conventional *Human-Centered Automation* idea that assumes superior priorities of a human over a machine, we put an emphasis on *increasing* interactions between them, accepting their *conflicts* and letting them cooperate with each other to change a current conflict status into a new coordinated one (Sawaragi, 1999). We are now extending the current system towards the one based upon this socially-centered automation idea.

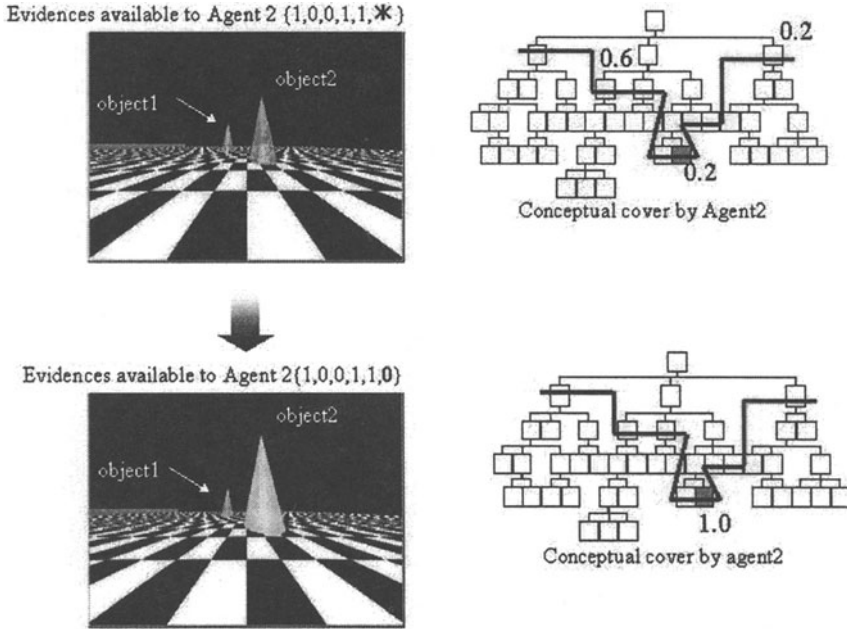


Figure 5 Pseudo-3D displays using VRML and corresponding conceptual covers

We have restricted what the agent can observe to the sensing information from the robot's environment. Another important source that is available to the agent is on what operations a human operator is taking. The observation of an operator's actions would enable the agent to infer his/her recognizing state of the environment, apart from the evidences that the robot can obtain from its task environment. Thus, the agent can reason about the internal states that a human operator grasps as well as about the true states that the agent itself infers from evidences from the robot's environment. As for the former, we can also construct another decision model that prescribes what kinds of recommendations to be given to a human operator by getting evidences from his/her behavioral features being mediated by inferred plausible hypothesis on the target objects. The conceptual covers constructed for those hypotheses do not always coincide with the one obtained for the latter, but may be differentiated, although their individual extension sets of objects are the same. Based upon the degrees of mismatch between the agent's and the operator's conceptual covers, interactions acted by the agent against the human operator can be altered and varied. Such a change of the agent's ways of interacting with the human

operator would alter his/her reactions, which would further make the agent's grasping states of the operator change to a new one. In this way, a *cyclic, mutually binding* relation does emerge, where both the human and the agent experience conflicts as well as agreements, and they can get to share respective concepts. This is a core of our proposing *socially-centered automation*.

8. CONCLUSIONS

In this paper, we put an emphasis on the fact that the agent should be embedded within the interactions with a human user, and for realizing this, establishing transparency of what the agent is grasping and how it is inferring is essentially important. We presented a situation awareness model for an agent integrating both of a normative decision theory and a machine learning method, which was applied to the interface design of a teleoperated mobile robot. Finally, we presented our perspectives towards realizing a socially-centered.

9. REFERENCES

1. Fisher, D. (1987). Knowledge Acquisition via Incremental Conceptual Clustering, *Machine Learning*, 2, pp.139-172.
2. Horvits, E. (1991). Time-Dependent Utility and Action under Uncertainty, *Proc. of the Seventh Conference on Uncertainty in Artificial Intelligence*, Los Angeles, pp.151-158.
3. Howard, R.A. and Matheson, J.E. (1983). Influence Diagrams, in Howard, R.A. and Matheson, J.E. (Eds.), *The Principles and Applications of Decision Analysis*, Strategic Decision Group, Menlo Park, CA.
4. Klein, G.A. et al. (1993). *Decision Making in Action: Models and Methods*, Ablex Pub. Corp., Norwood, NJ.
5. Klein, G.A. (1993). A Recognition-Primed Decision Model of Rapid Decision Making, in Klein, G.A. et al. (eds.), *Decision Making in Action: Models and Methods*, pp.138-147, Ablex Pub. Corp., Norwood, NJ.
6. Maes, P. and Kozierok, R. (1993). Learning Interface Agents, *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pp.459-465.
7. Maes, P. (1994). Agents that Reduce Work and Information Overload, *Communications of the ACM*, 37-7, pp.30-40.
8. Poh, K.L., Fehling, M.R. and Horvitz, E.J. (1994). Dynamic Construction and Refinement of Utility-Based Categorization Model, *IEEE Trans. of System, Man, and Cybernetics*, 24-11, pp.1653-1663.
9. Rouse, W.B. (1988). The Human Role in Advanced Manufacturing Systems, in Compton, D. (Ed.), *Design and Analysis of Integrated Manufacturing Systems*, National Academy Press, Washington, D.C..
10. Sawaragi, T., Iwai, S., Katai, O. and Fehling, M.R. (1994). Dynamic Decision-Model Construction by Conceptual Clustering, *Proc. of the Second World Congress on Expert Systems*, pp.376-384, Lisbon, Portugal.
11. Sawaragi, T., Takada, Y., Katai, O. and Iwai, S. (1996). Realtime Decision Support System for Plant Operators Using Concept Formation Method, *Preprints of International Federation of Automatic Control (IFAC) 13th World Congress*, Vol.L, pp.373-378, San Francisco.
12. Sawaragi, T., Katai, O. (1999). Bio-Informatic Activity Modeling for Human-Artifacts Symbiosis under Resource Boundedness, *Artificial Life and Robotics*, 3, pp.45-53.
13. Sawaragi, T. (1999). Modeling and Analysis of Human Interactions with and within Complex Systems, *Proc. of IEEE Conference on System, Man, and Cybernetics*, Tokyo.
14. Sheridan, T.B. (1997). Human-Centered Automation: Oxymoron or Common Sense?, *Proc. of IEEE Int. Conf. on System, Man and Cybernetics*, Vancouver, Canada.