Intelligent Systems Validation Tool based on bi-directional many-sided explanation typed multi-step Validation Method

Setsuo Tsuruta* and Takashi Onoyama**

* Systems Development Laboratory, Hitachi, Ltd.
1099 Ohzenji Asao-ku Kawasaki-shi, 215-0013, Japan
e-mail: tsuruta@sdl.hitachi.co.jp
** Research & Development Department
Hitachi Software Engineering Co., Ltd.
6-81,Onoe-cho,Naka-ku,Yokohama,231-0015,Japan
e-mail: onoyama@ori.hitachi-sk.co.jp

Abstract

A bi-directional many-sided explanation typed multi-step validation method including its implementation architecture has been proposed to diminish validation loads for busy experts. This paper presents a validation tool based on this method. This tool supports the multi-step, bi-directional validation by experts, KEs (Knowledge Engineers) and computers on the Intra/Internet environment. Particularly this tool anonymously supports efficient validation and its process management for preventing validation delays of busy experts. Our experimental evaluation indicated that this tool could diminish experts' validation loads to less than 20% compared to when the tool not employed.

1. Introduction

As to validation methods for intelligent systems, validation methods based on Turing tests (Turing 1950; Knauf et al. 1998; Knauf, Gonzalez and Jantke 1999) have been proposed. Though these are effective theories and ideas for validation of expert systems, they have an actual problem that they impose too much validation loads and responsibility upon busy experts. A validation method using the Internet (Jantke and Stephan 1998) and a Turing-Test-like tool through a network in order for experts to anonymously cooperate in validation tasks have also been proposed (Salecker and Knauf 1999). Nevertheless, these methods do not diminish the foregoing validation load problems. Experts' real intention is that they do not want to cooperate. This is because they are in such a delicate social position as to lose their vocational superiority with very high probability, if high-level intelligent solving methods become available on the market as products (Tsuruta et al. 1997).

To solve such problems, a bi-directional explanation type multi-step validation method (Tsuruta et al. 2000) and its implementation method (Onoyama et al. 2000) were proposed. This paper presents the validation tool based on these validation methods. Using this tool, experts effectively and anonymously shares validation loads with KEs and computers. Thus, experts' load on validating intelligent systems validation tasks can be lessened and the difficulties of experts' cooperation in validation tasks can be smoothed away.

Section 2 describes the previously proposed validation method and the problems in developing the tool based on this method. In section 3, the newly proposed tool and its human interface are described. In section 4, the validation tool is evaluated and in section 5, the tool is compared with others and the results are concluded.

2. Validation Method and its Problem

2.1 Validation method for intelligent systems

Intelligent system's validation basically needs experts' cooperation, but black-box style tests based on the Turing Test cause the heavy validation load on experts. This heavy validation load on experts prevents accomplishment of intelligent system's validation.

In order to solve this problem, a concept called "bi-directional many-sided explanation typed multi-step validation method" was proposed (Tsuruta et al. 2000). This concept has the following three characteristics:
1) Multi-step validation in three steps, namely, validation by computers, validation by KEs (Knowledge Engineers) and validation by experts. 2) Many-sided explanation types to multilaterally edit and explain test results of a solving method ant its component methods in each profile of the process in forms required at each of the foregoing these steps. These forms are statistical data, charts, graphs, topological maps and road maps. 3) Bi-directional interaction under a distributed information environment in each of these steps using these explanations.

2.2 Implementation architecture

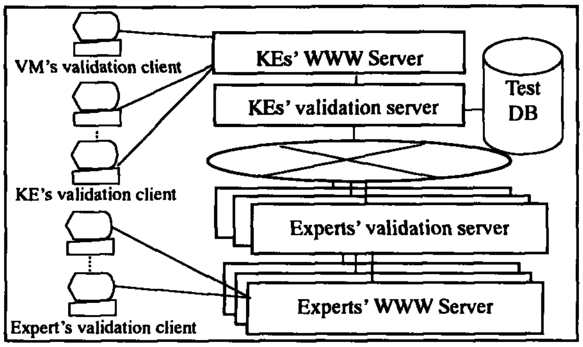


Figure 1. Architecture for the proposed validation method

The basic architecture for implementing the above validation method has already been proposed (Onoyama et al. 2000). However more anonymous but efficient validation is required.

This architecture is composed of validation servers, WWW servers, and validation clients as shown in fig.1. The validation servers consist of KEs' validation servers and experts' validation servers. In order to maintain experts' anonymity, the experts' validation servers are distributed to each expert's working place and are connected with KEs' validation servers through the Intra/Internet. Each validation server is connected with WWW servers respectively. The validation clients consist of KEs' validation clients and experts' validation clients . KEs (including the validation manager: VM) use KEs' validation clients. Each validation client accesses the WWW servers by way of the Web browser. The KEs' validation server is connected to the KEs' WWW server. And, each experts' validation server is connected to each experts' WWW server, respectively.

The validation environment under this architecture is as follows: 1) Multiple experts, multiple KEs and a VM (possibly having assistants) can participate in the validation. 2) Each expert possibly belongs to different sections, and carries out his work, including validation, away from KEs and the VM.

The validation process (Onoyama et al. 2000) consists of the 6 steps such as Test case & plan generation, automatic validation, validation by KEs, validation by experts, aggregation, and process management as shown in fig. 2.

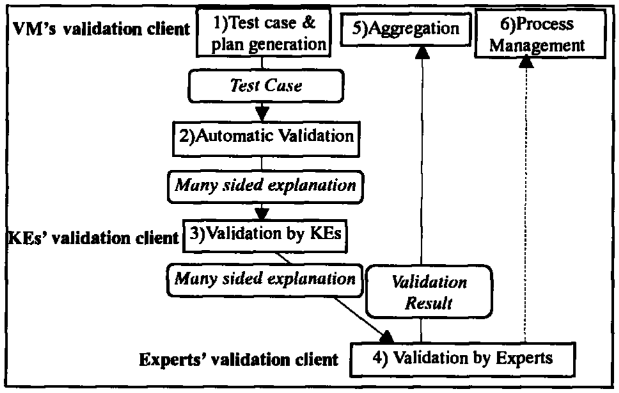


Figure 2. Validation Process

2.3 Problems in developing validation tool

The following problems became evident in developing the validation tool based on the above validation method.

1) Lack of the process management

Since experts are usually very busy and placed in delicate social positions, there are few motivations that experts voluntarily cooperate for validation. Yet, serious problems such as delays in their validation process are frequently not noticed at all, as they work anonymously and away from the VM and KEs. Therefore, experts' validation process needs to be monitored for such delays, which should be managed to be recovered. Thus, it is necessary for the VM or KEs to efficiently manage experts' validation process.

2) Lack of more anonymous and efficient validation

Since experts should never be disgraced because of their delicate social position, the final responsibility must not be put on experts and each expert should participate in the validation process anonymously. To keep this anonymity of experts, requests for validation to experts and the reception of their validation results are carried out by way of the Intra/Internet environment.

However, more anonymous but efficient validation is required. This reason is that experts are busy and efficiency is important. Furthermore lack of anonymity will possibly cause expert to weaknesses being exposed, their occupational judgment being interfered and their vocational superiority being undermined. In such case it can be easily imagined that the expert placed in the above mentioned delicate position would apparently reject or become reluctant to cooperate in subsequent validation.

3) Lack of human interface

The following human interface to meet the above requirements is lacking.

a) User interface for process management

The VM and KEs have to watch each expert's validation status to advance validation tasks without delay. They are also expected to know not only each expert's validation process, but also the situation of each expert's daily work (e.g. in terms of quantities, transfers, holidays, etc.) Otherwise, they are not able to assign validation tasks to experts. Since experts participate in validation tasks anonymously, the VM (and KEs) cannot directly ask questions to the experts. Even the VM should not know the experts' real names, as their anonymity have to be perfectly assured.

Thus, a user interface that enables a VM to grasp the each expert's validation process without knowing each expert's name is required.

b) Interface for many-sided explanations and bi-directional interactions

Busy experts manage to carry out validation tasks in parallel with their daily works. Thus, efficient access to many-sided explanations which contain different graphs, tables, and road maps for the experts' benefit, is required as a tool of the interface. Yet, to realize responsible validation, experts need to ask questions to KEs and have answers from them through bi-directional interactions. These should naturally be done anonymously. It is also necessary that experts ask questions about many-sided explanations provided.

3. Validation Tool and Human Interface of the Proposed Tool

3.1 Concept of the validation tool

A validation tool is proposed to solve the above problems. The concept of this tool is as follows:

1) Guarantee for experts' anonymity

To keep experts' anonymity, the discrimination of experts is done only through IDs. Yet, to give them more assurances of their anonymity, two kinds of IDs (namely, ID for KE and ID for expert) are provided, the latter is used by experts and the former is used by KEs respectively. A conversion table between these two kinds of IDs is encrypted and stored in the KEs' validation server. And another conversion table between IDs for experts and experts' real names (e-mail addresses) is also encrypted and stored in the experts' validation server.

A user interface is provided for experts' validation client. This interface enables experts to input IDs, as well as enables experts to store encrypt key in experts' validation servers. Since each expert inputs his ID and his encrypt key, KEs and a VM are not able to discover experts' real names.

2) Process management under anonymous environment

To keep experts' anonymity, the following process management using only the above mentioned IDs for discrimination of experts is realized. First, for each test case to be stored in Test DB, ID for discrimination of an expert assigned to each test case and the validation processing status (not started/ under validation/ finished) is recorded. When an expert starts validation tasks on his experts' validation client, or when he finishes his validation tasks, the validation status is sent to the KEs' validation server through the experts' validation server. The KEs' validation server modifies the process status of the test case stored in the Test DB when the server received new information. The KEs' validation server aggregates the process status of the test case to be provided for the VM.

A user interface for experts to input the schedule of their work other than validation is also provided to the experts. This information is also sent to the KEs' validation server to be provided for KEs.

3) Bi-directional interaction corresponding to many-sided explanations

An input area is provided to the display screen for many-sided explanations, so that experts can ask KEs about these explanations. Yet, in this user interface (we called it "many-sided explanation typed user interface"), replies from KEs are also shown on the same display screen. Experts' questions about many-sided explanations and KEs' replies to these questions are stored in the Test DB. In the Test DB these questions and answers are all linked with their related many-sided explanations. That is, many-sided explanations have links to experts' questions. Experts' questions have links to KEs' replies. As well, KEs' replies have link to experts' questions to these replies. And this chain of links is continued.

3.2 Functions of the proposed validation tool and its human interface

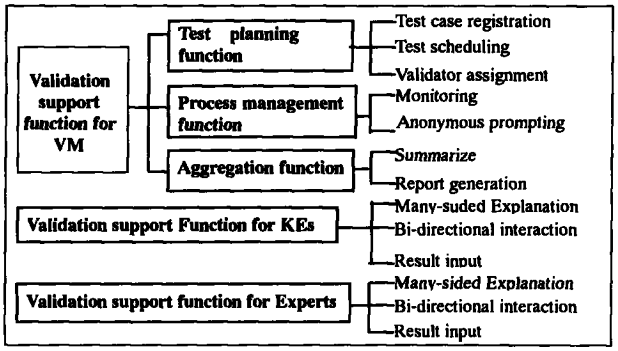


Figure 3. Function of the proposed validation tool

To realize the above concept the proposed tool has the following functions as shown in fig.3. The VM, KEs and experts use the functions by way of Web browser. Hereafter, screens in validating the delivery scheduling intelligent system are used for examples (fig. 4,5).

3.2.1 Validation support functions for VM

These functions (many-sided explanation typed user interface for KEs) are provided in a VM's validation client and support VM's test planning, validation process management and aggregation of the validation result.

1) Test planning function

a) Test case registration functions

This function supports to store generated test cases and register to the Test DB.

b) Test scheduling support function

This function enables a VM to assign a KE responsible for each test case and to set its deadline.

c) Validator assignment function

Through this function, a VM assigns a validator, namely experts in charge of validation when the KE needs the validation by experts. To keep experts' anonymity, the VM discriminates each expert only by his ID. Therefore, the IDs of the experts, whose validation is on schedule, are displayed on the VM's validation client. Yet, this function checks the number of delays in the validation of each expert. When the number of delayed test cases of an expert reaches to the specified value, his test cases will be also assigned to another expert ahead his schedule.

2) Process management function

a) Monitoring function

Since experts are already heavily loaded with their daily work, their cooperation cannot be readily obtained. What is worse, their real intention is that they do not want to cooperate on validation. This is because they are in such a delicate social position as to lose their vocational superiority, if intelligent systems become available on the market as products. Thus, their validation often delays.

Therefore, this function of the proposed validation tool monitors the experts' validation processes, and if the validation deadline is over, an alarm (namely, the number of delays and its graph) is indicated on the screen as shown in fig. 4. This process management function displays only IDs that can discriminate experts and does not display their real names (or their real e-mail addresses) to keep experts' anonymity.

b) Anonymous prompting function

If there are any delays in experts' validation, this function enables the VM to specify an ID for discriminating an expert and can ask him anonymously to urge his validation tasks by sending messages using this ID. Or else this enables the VM to ask other experts ahead schedule to take over the validation.

3) Aggregation function

This function summarizes validation results of automatic, KEs' and experts' validation stored in the Test DB and generates a validation result report.

3.2.2 Validation support functions KEs

These functions are provided in the KEs' validation client, and supports KEs to validate and to answer questions of experts.

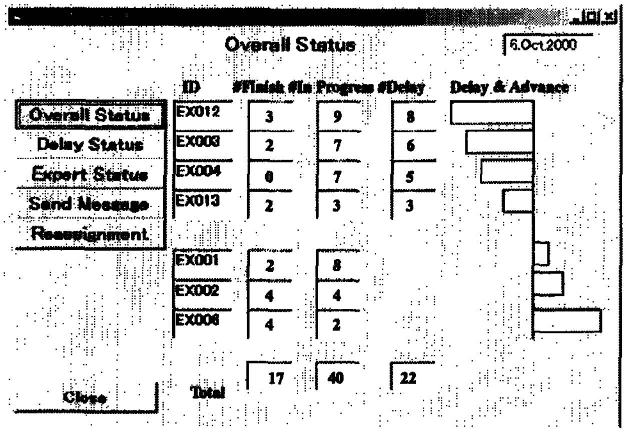
1) Display function of many-sided explanations

These functions display many-sided explanations such as topological maps, statistical values, graphs and charts for the validation by KEs. Fig. 5 is an example display screen in the validation of a delivery scheduling intelligent system.

Comment input area is added to the display screen of many-sided explanations. Experts and the VM can refer to the comments inputted in the area by KEs.

2) Support function for bi-directional interaction

This function enables KEs to perform the bi-directional interactions (questions and answers) with experts, anonymously and efficiently. As to the efficiency, KEs questions to experts are displayed on the experts' validation client immediately upon its arrival for notifying it. Meanwhile, for the prompt response of KEs (to experts), requesting message for questions (from experts) is displayed on the KEs' validation client upon arrival.



Overall Status

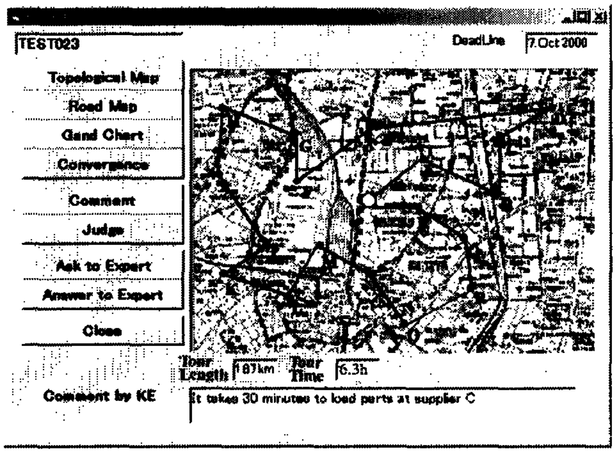
6.Oct.2000

ID	#Finish	#In Progress	#Delay	Delay & Advance
EX012	3	9	8	
EX003	2	7	6	
EX004	0	7	5	
EX013	2	3	3	
EX001	2	8		
EX002	4	4		
EX006	4	2		
Total	17	40	22	

Figure 4. Monitoring screen

3.2.3 Validation support functions for experts

The experts' validation client has the following functions (many-sided explanation typed user interface for experts), which supports validation by experts: Display function of many-sided information, bi-directional interaction function for experts to ask questions and information to KEs. And inputting function of validation results by experts.



TEST023

DeadLine 7.Oct 2000

Tour Length 187km Tour Time 6.3h

Comment by KE: It takes 30 minutes to load parts at supplier C

Figure 5. Many-sided explanation screen

1) Many-sided explanation function

On the experts' validation client, many-sided explanations necessary for experts' validation such as road maps and charts are displayed. In this display screen, comments added by KEs are shown together with the many-sided explanation.

2) Bi-directional interaction function

When experts need to ask questions or more detailed explanations, they can input their requirements into the question input area in the lower part of the display screen of the many-sided explanations. These inputted questions and requirements are sent to the KEs' validation client.

In order for busy experts to shorten communication time with KEs, a request-prompting message is displayed on the KEs' validation client upon the arrival of experts' asking. As well, prompts to notify experts of KEs' responses are displayed on the experts' validation client, immediately upon the responses' arrival.

3) Validation results input function

In order for busy experts to readily input their validation results, the validation results of experts can be input as an evaluation value which is a digit of 5 grades. The grade 1 means "reject". The grade 2 through 5 means "pass". Experts intuitively input an evaluation value and add a short comment in natural language, using this user interface.

4. Evaluation

1) Effect of multi-step validation

Our tool was applied for validating a delivery scheduling intelligent system, and was evaluated. For one validation experiment, 20000 test cases were used and only 300 cases out of 20000 test cases, which are 1.5% of the total test cases, were needed experts' validation. And in another experiment, around 20 % test cases (180 cases out of 1000 test cases) were needed experts' validation.

2) Effect of many-sided explanation typed user interface
For the examination of these 20000 test cases, three KEs used to take 4 weeks. However, using the many-sided explanation typed user interface they could shorten the time to 2 weeks.

Further, validation of 100 test cases out of the above 300 test cases which needed expert' validation was validated by two groups of experts, respectively. One of the groups validated using our tool but without a VM. The other group validated without using our tool. The former group finished their validation within 1.4 months, but the latter group took 3 months. Thus, the many-sided explanation typed user interface of our validation tool shortens the time of validation to 50%.

3) Effect of process management
Another 100 test cases out of the above 300 test cases were validated by two groups of experts, where only the deadline of validation tasks was advised to one of the two groups though both used our tool. Meanwhile, the daily progress in the validation tasks of another group was monitored and the daily assignment of validation tasks was adjusted so that the number of test cases in progress (under examination) to be below 10 every day. As a result, the former group could not finish their tasks within the deadline (1 month), but took 1.5 months. On the other hand, the latter group could finish their tasks within 1 month. Thus, the validation process management could shorten the validation time to 2/3.

4) Effect of the reassignment
In the validation process of still another 100 test cases, there were 15 delays without the above daily adjustment of validation tasks' assignment. But the VM could immediately know all these delays, using our tool. Accordingly, he could promptly reassign the delayed test cases to other experts considering the number of test cases under validation (in progress), the amount of each experts' tasks except validation, and historical and/or statistics data of each expert's delay, all of which can be displayed in the monitoring screen of his validation client PC. Due to such reassignment of test cases, validation of 100 test cases was completed within one month as is initially scheduled.

5) Effect of bi-directional interaction user interface
In the validation process of first half 150 test cases of the above-mentioned 300 test cases, experts made 130 questions to KEs. The KEs responded to these questions by using many-sided explanations from their validation client PCs. Thus, experts and KEs interacted with each other by using the user interface of the validation tool. And their questions and answers linked with test cases were automatically recorded in the Test DB.

In the validation process of the latter half 150 test cases, KEs prepared comments of test cases for experts by referring to questions and answers previously recorded in the Test DB. These comments contributed to reduce experts' questions to only 40 in validation process of the latter 150 test cases.

5. Comparisons and Conclusion

Many excellent theories and ideas based on the well-known Turing Test have been proposed as methods to validate AI systems (Abel and Gonzalez 1997; Knauf and Gonzalez 1997; Knauf et al. 1998; Knauf et al. 1999; Salecker and Knauf 1999). Turing-Test-like tools through a network in order for experts to cooperate in validation tasks have also been proposed (Salecker and Knauf 1999). However, all these methods and tools do not propose methods or mechanisms to diminish the foregoing validation load problems such as validation loads sharing or efficient and many-sided explanation typed user interface for busy experts. Furthermore, the management method of validation process of experts, who are very busy and placed in a delicate social position, is not proposed.

Using our proposed tool supporting multi-step validation and bi-directional but anonymous many-sided explanation typed user interface, experts can efficiently and anonymously validate intelligent systems, with their validation loads shared by KEs and computers. Thus, experts' validation tasks can be lessened and difficulties of experts' cooperation concerning with validation are smoothed away.

And, using our tool the VM or KEs can manage the process of validation tasks of each expert on the Inter/Intra network. This prevents delays in experts' validation tasks.

References

Abel, T.; and Gonzalez, A.J. 1997. Enlarging a Second Bottleneck: A Criteria-based Approach to Manage Expert System Validation Based on Test Cases Procs. FLAIRS'97

Janke, K.P.; and Stephan. A. 1998. Scenarios of Remote Validation in the Internet, MEME-MMM-98-5.

Knauf, R., Jantke, K. P., Abel, T., and Philippow, I. 1997. Fundamentals of a TURING Test Approach to Validation of AI Systems. Proc. IWK'97.

Knauf, R.; Jantke, K. P.; Abel, T., and Philippow, I. 1998. Fundamentals of a TURING Test Approach to Validation of AI Systems. Proc. IWK'98.

Knauf, R.; Gonzalez, A.J.; and Jantke, K. 1999. Towards Validation of Case-Based Systems. Proc. FLAIRS'99.

Knauf, R.; and Gonzalez, A.J. 1997. Estimating an AI system's validity by a TURING Test. Proc. IWK'97.

Onoyama, T.; Kubota, S.; Oyanagi, K; and Tsuruta, S. 2000: Implementation Method of a Validation tool for Intelligent Systems, Proc. IWK2000.

Salecker, D.; and Knauf, R.1999. Validation Manager A Tool for AI System's Evaluation by A Turing Test-Like Methodology. Proc. FLAIRS'99.

Tsuruta, S.; Eguchi, T.; Yanai, S.; and Ooshima, T. 1997. Dynamic Goal centered Coordination AI system: Application and the validation problem. Proc. IWK'97.

Tsuruta, S.; Onoyama, T.; Kubota, S.; and Oyanagi, K. 2000. Validation Method for Intelligent Systems. Proc. FLAIRS2000.

Turing, A. M. 1950. Computing machinery and intelligence, Mind, 59(236):433-460.