A Simulation-based Procedure for Expert System Evaluation

Yang, Chunsheng; Kose, K.; Phan, Sieu; Kuo, P.

National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de Technologie de l'information

NRC-CNRC

*A Simulation-based Procedure for Expert System Evaluation**

C. Yang, K. Kose, S. Phan, and P. Kuo
December 1999

***published in** Proceedings of the IEA/AIE 13th International Conference, New Orleans. June 19-22, 2000. 15 pages. NRC 43621.

A Simulation-based Procedure for Expert System Evaluation

CHUNSHENG YANG[1], KUNIJI KOSE[2], SIEU PHAN[1], AND PIKUEI KUO[3]

[1] *National Research Council, Montreal Rd, M-50, Ottawa, ON, K1A 0R6, Canada*

{Chunsheng.Yang, Sieu.Phan}@iit.nrc.ca

[2] *Faculty of Engineering, Hiroshima University, Higashi-Hiroshima, Japan*

[3] *National Taiwan Ocean University, Keelung, Taiwan, Republic of China*

pkkuo@mail.ntou.edu.tw

Abstract

It is well recognized that the evaluation of the knowledge-based system is very important and difficult in the development of expert systems in the domains such as aviation and navigation. To alleviate some of the difficulties, and to reduce the development cost, the authors proposed an evaluating procedure based on simulation. The proposed procedure was designed to validate the functionality and capability of an expert system that was developed to provide decision-making support in ship navigation. In this paper, the developed expert system for collision avoidance is first outlined, and the procedure to evaluate the developed expert system by using a simulation-based approach is presented in details. It is also concluded that the simulation-based procedure is feasible and effective for evaluating expert systems in a number of different domains.

1 Introduction

In the development of real-world problem expert systems such as collision avoidance in ship navigation or intelligent alarm correlation in telecommunication, the evaluation of the developed knowledge-based system is one of the most important stages. Before a knowledge-based system can be deployed, it must be evaluated for accuracy. Generally, the knowledge-based systems designed to solve real-world problems are very large and contain thousands of rules; thus verification becomes difficult. First, it is hard to examine possible interactions of rules simultaneously. Second, in order to carry out verification checks, it is necessary to obtain certain additional information such as observable variables. Third, it is difficult to set criteria for each observable variable. There have been lots of research in the field of verification and validation (V&V) of the knowledge-based systems. There have been many achievements in this field. These results focus on V&V theory and techniques [1][2][3][5], V&V systems and specifications, and V&V applications [4][6][7].

Some examples are as follows: M. Benerecett [1] applied model checking technique to multiagent system verification; D. Fensel [7] deployed KIV (Karlsruhe Interactive Verifier) for the verification of conceptual and formal specifications of knowledge-based systems; and M. Ramaswamy [4] presented a technique based on directed hypergraphs that enables developer to determine overall integrity of the rule bases by verifying partitions locally. All of these techniques are very useful for evaluating the knowledge-based systems from the viewpoint of accuracy of knowledge modeling and rule bases. However, it is expected that evaluation of knowledge-based systems can be carried out in a realistic environment, which allows domain expert to evaluate the knowledge-based systems directly. It is impossible to directly test and evaluate the developed expert system in a real application, because the procedure is too costly and unsafe, especially in the domain of aviation and navigation. To reduce the cost of development and perform evaluation in a realistic environment, simulation is considered to be the most effective and feasible approach, especially for knowledge-based systems in the ship navigation and aviation domain. The authors suggest two kinds of simulation: fast-time simulation and real-time simulation. Fast-time simulation is a scenario-based approach to test the knowledge-based systems. It is used to test the common basic problem-solving ability of the knowledge-based systems and to fine-tune the knowledge bases. Real-time simulation is a realistic environment, which can give the human being a real-world feeling and allow domain expert to evaluate the problem-solving ability of the expert system directly. Based on such simulations we proposed an evaluating procedure for domain-oriented expert system such as collision avoidance expert system. The evaluating procedure includes four steps: identifying evaluation purposes, identifying evaluation items and indices, designing and performing simulation tests, and analyzing results. In this study, the authors developed a collision avoidance expert system to assist navigators in the decision-making process in ship navigation [8][9][10]. Then we evaluated the developed system based on the proposed evaluating procedure. In this paper, we outline the developed expert system for collision avoidance in Section 2; then the simulation-based evaluating procedure is presented in Section 3; and how to evaluate the developed expert system using the proposed evaluating procedure will be discussed in Section 4. Conclusions are given in the last section.

2 Developed Expert System for Collision Avoidanc

The goal of the collision avoidance expert system is to assist ship navigators in their decision-making process to avoid collision. The developed knowledge-based system has been incorporated into the Integrated Navigation System (INS) [8][9] at Hiroshima University as an intelligent decision-making support subsystem [10]. In this section, the outline of the developed expert system is presented in order to describe how to evaluate it in later sections.

It is well known that the procedure followed by a captain during the ship handling to avoid collision consists of collecting information, assessing the encounter situation, determining the collision avoidance action and executing the action. In making a decision for avoiding collision, the captain decides the action of the collision avoidance using encounter situation, traffic regulations, his experience, and judgment obtained by visual information. The feasibility of such a decision is limited by the constraints of the visual field, misunderstanding of the information and the action of the target ships, executing miss of the actions, and so on. When the captain decides to avoid a collision, he can only focus his attention on the most dangerous target, and cannot pay attention to other target ships encountering with his ship due to his capability of information processing. Therefore, the authors proposed to develop a knowledge-based collision avoidance support system, which is able to provide decision-making support for operators. In other words, such a support system should be able to compensate the human deficiency and to provide an effective maneuvering action to operators. Such an expert system should possess the following abilities:

- Sophisticated problem-solving ability;
- Full safe navigating ability;
- Prediction of target ships` action ability;
- Interactive ability; and
- Real-time responding ability.

To reach these objectives, the authors concentrate on the development of a knowledge-based system for identifying an effective collision avoidance action. To analyze effectively multi-ship encounter situations, the authors introduce a target ship classification method and the concept of the most dangerous ship, the dangerous ship, the restricting ship, the indifferent ship, and the unblocking scope of ship handling space. To improve the feasibility and safety of the collision avoidance action, the authors propose a predicting method called a prediction of plural action for the action prediction of target ships. Figure 1 shows the inference sequence of the system. In terms of the encounter situation and the safety evaluation, the inference engine selects necessary sequence for the inference procedure. The inference procedure uses knowledge bases built for the developed expert system. As shown in Figure 2 the collision avoidance expert system is designed based on hierarchical architecture and modularized knowledge structure. The top layer in the system is the inference control. It is responsible for the control of the inference procedure. The second layer is the main knowledge bases, which include the classification of target ships, the prediction of target ships' action, the identification of the method of collision avoidance, and the establishment of course-line waypoints. The third layer contains knowledge modules of every knowledge base. The fourth layer includes the preliminary knowledge modules such as traffic regulation, identification of target ships, and so on. The system consists of the following inference sequences:

(1) Prediction of target ship's scheduled action

First, the system uses a knowledge base to classify the navigation environment into one of three categories: open sea, coastal, or route navigation. The prediction of the target-scheduled action depends on this classification: in the case of open sea or coastal navigation the target will maintain its current course and speed, and in the case of the route navigation, it will follow the route

(2) Classification of target ships

After the computation of collision risk of target ships using of the Nagasawa Risk Model [11], the target ships are classified as the most dangerous , the dangerous, the restricting, or the indifferent ship depending on their risk of collision. A dangerous ship is defined as a ship having risk of collision that exceeds the safe level when both ships maintain their scheduled course lines. In the case of several dangerous ships, the most dangerous ship has the highest risk among dangerous

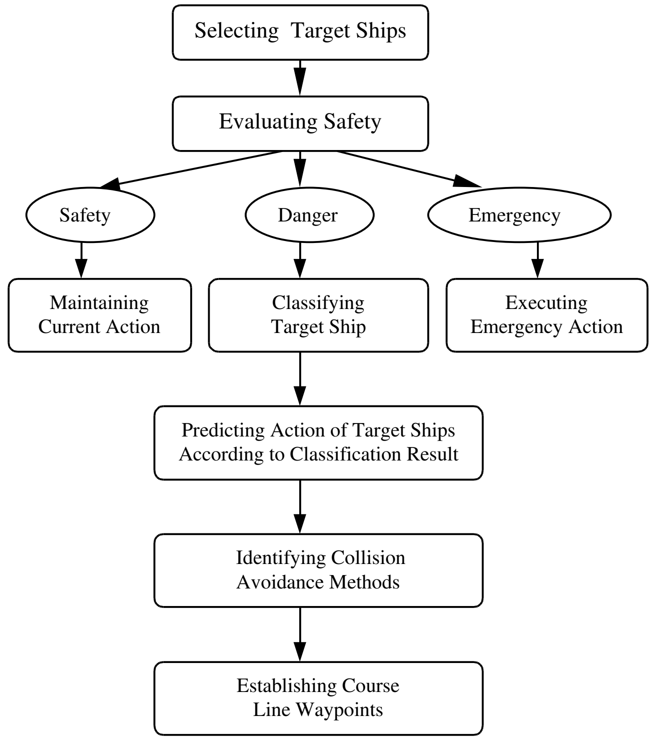

Fig.1 The Inference Sequence of the System

ships. A restricting ship is defined as a ship that will cause no danger if own ship and target ship maintain their scheduled course lines, but it will frustrate the action of own ship if she takes the collision avoidance action for a dangerous ship. The target ship that lies outside the maneuvering space of the own ship is defined as the indifferent ship.

(3) Prediction of target ship's collision avoidance action according to the classification of the ship

For the dangerous ships and restricting ships, their collision avoidance actions are predicted by using the same approach and knowledge base as that of the own ship. The predicted actions of collision avoidance of target ships will be incorporated into the procedure when own ship's action of collision avoidance is formulated.

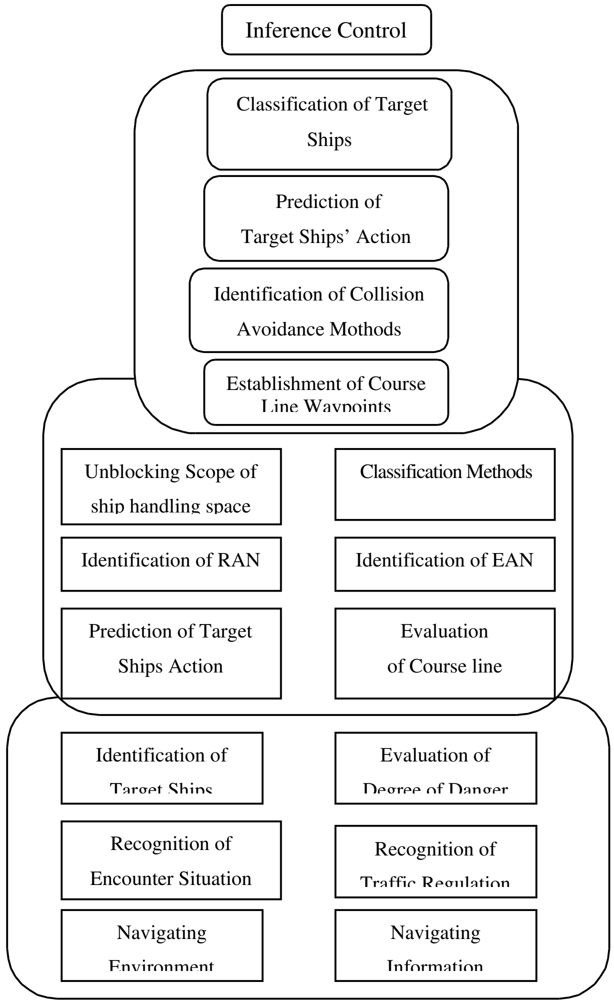

Fig.2 The Architecture of the Developed Collision Avoidance Exert System

(4) Establishment of the course line waypoints as the collision avoidance action

The action of collision avoidance of own ship is formulated basically against the most dangerous ship. The action in the own ship maneuvering space is evaluated considering the prediction of the action of the dangerous ships and the restricting ships. As a result, the most efficient and feasible action is selected.

3 Simulation-based Evaluating Procedure

After we built the prototype of the expert system by using Nexpert Object[1], C Language and Unix Platform (SUN Workstation), we have to test and evaluate the developed expert system, and show the system's capability to the domain expert. Otherwise, it is impossible to deploy the developed system to the real application field. Therefore, how to test and evaluate the expert system becomes very important. Because different domain has different requirements, different application background, different technical support, and different targets, it is very difficult to propose the same approach and criteria for evaluating different domain expert systems. Of course, one might argue that the best way to test the system is to try it out in a real application environment. However, it will largely increase the cost and cause unexpected thing happen. It is impossible for a domain operator to accept such an approach. So, we proposed a simulation-based procedure for evaluating the expert system. To systematically and effectively evaluate the expert system, we suggested that the simulation-based approach should include the following procedures:

- **Identifying evaluation purposes**

  In terms of the requirements and the goals of the development, we need to identify the evaluation purposes for the expert system.

- **Identifying evaluation items and their indices**

  To meet the above evaluation purposess, we need to determine the evaluation items that can reflect the original design requirements and evaluation purposes. In order to prove that the evaluation items are feasible and acceptable, for each evaluation item, we need to specify some indices. Such indices might be either quantitative or qualitative.

[1] Nexpert Object is an expert system development environment which can provide production knowledge representation and object knowledge representation. It can also provide a powerful GUI and API for developers to develop knowledge-based system.

- **Designing and performing evaluation simulation**

  Some of the above evaluation indices can be obtained by personal judgment from the designers or domain experts; some can be obtained from simulation. We suggested that it is better to carry out two kinds of simulations: fast-time simulation and real-time simulation. Fast-time simulation can be used to test typical ability of the expert system, to quickly identify problems in the knowledge base, and to fine-tune the knowledge base. Real-time simulation can be used to provide a realistic environment for testing the system. In real-time simulation, we need to ask the domain experts to evaluate the system and obtain their comments and feeling about the system, then to improve the system. Therefore, to carry out such simulation, it is necessary to design the simulation scenario, to develop simulation environment that allows operators to interact with the knowledge base.

- **Analyzing the results**

  As described above, evaluating indices might be either quantitative or qualitative. Therefore, after carrying out the evaluating simulation, we need to figure out these indices for judging the evaluating items by analyzing the simulation results quantitatively or qualitatively.

Corresponding to different domain expert systems, the specific content of the above procedure might be different. Therefore, let us take our developed expert system as an example to describe how to evaluate expert systems using the proposed evaluating procedure.

4 Evaluation of the Developed Expert System

In this section, how to evaluate the developed expert system for collision avoidance decision-making support by using the proposed simulation-based evaluating procedure is discussed.

4.1 Evaluation Purposes

According to the requirements and developing target of our domain expert system, the evaluation purposes are to validate the capability of decision-making support, to test the problem-solving ability of collision avoidance, and to identify the problems for further refinement of the knowledge base.

4.2 Evaluation Items and Indices

To meet the above objectives, we define the following items as the evaluating items in terms of the designing requirements and the targets of the developed system:

- collision avoidance prolem-solving ability;
- ship navigation safety; and
- decision-making support ability.

The collision avoidance problem-solving ability is defined as the fact that no collision will happen during ship navigation and the collision avoidance action is reasonable and does not disturb other ships. The ship navigation safety is defined as the fact that the desired safety level must be satisfied during ship navigation. And the decision-making support ability is defined as the fact that the system should be able to provide the effective support for operator's decision-making and to alleviate their burden on decision-making. In terms of these items, we define some indices that may be either quantitative or qualitative. The quantitative indices are:

- the safety level of own ship,
- the target ship risk related to own ship action,
- the deviation of course line, and
- the response time,

and the qualitative indices are:

- the readable information of decision-making support,
- the feasibility of proposed action, and
- the accuracy of the collision avoidance action.

All of these indices have to be obtained from fast-time simulation or real-time simulation. For qualitative indices, we asked domain experts to give their evaluation by using questionnaire. For the quantitative indices, we derived the results from the simulating results by analyzing collected traffic data. The safety level of own ship is calculated following the Nagasawa's risk model [11], and the target ship risk related to own ship' collision avoidance action is defined as the risk level from the viewpoint of target ship due to the own ship's collision avoidance action. It is also calculated with the Nagasawa's risk model. The deviation of course line is defined as the distance between the scheduled course line and the course line of collision avoidance action. It reflects the cost of collision avoidance action.

4.3 Evaluation Simulation

To obtain the evaluation result of the above indices, two kinds of simulations are carried out. They are fast-time traffic simulation and real-time ship handling simulation.

(1) Fast-time Traffic Simulation

Traffic simulation is carried out for testing basic problem-solving ability. We ask navigators to build typical encounter situation scenarios according to their navigating experience and international traffic regulations. Using these traffic scenarios [9], the simulation is done in the developed traffic simulation environment as shown in Figure 3.

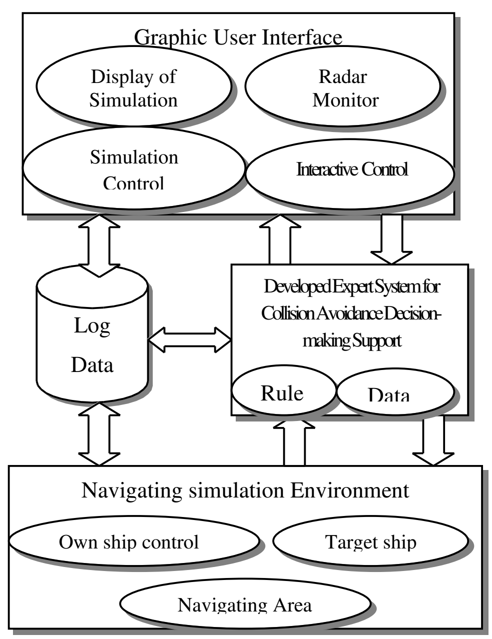

Fig.3 The Composition of Traffic Simulation System

This simulation environment includes navigating environment system, graphic user interface (GUI), the developed expert system, and simulation result log system. The navigating environment system can provide all of the necessary data for own ship, target ships, and navigating area (open sea, congested waterway, and route). The GUI shows simulation results which are ship trajectories with time history and allows the operator to interact with knowledge base system. The developed expert system is one that we want to evaluate; the log system can record history data of simulation and can reprint these recorded data to check the executing result of the knowledge-based system. These simulation data will be analyzed in the final analysis step.

(2) Real-time Ship-Handling Simulation

To carry out real-time ship-handling simulation, it is necessary to build a realistic simulation environment and simulation scenario.

Using the ship-handling simulator and the INS [9], we constructed a simulation system for the developed knowledge-based decision-making support system. This system is shown in Figure 4. Such a simulation system could provide a real-time ship handling environment. It possesses the following features.

- ship motion model: MMG model;
- ship guidance method: Optimum control;
- ship position system and its error: GPS and GPS error;
- reproducible navigating environment and scene of ship handling; and
- real feeling of the danger during ship handling of collision avoidance.

Such a simulation environment is very useful for evaluating the developed expert system, because ship-handling simulator could provide the scene of ship handling and the feeling of the danger during the collision avoidance to the operators. The developed knowledge-based system is incorporated into the INS as a decision-making support system. It might propose an action of collision avoidance to operators. Operators decide the final action according to their judgement with the help of support function of the INS. Meanwhile, the domain experts might easily evaluate the result of collision avoidance and give us their comments and requirement and feeling on the system. Simulation scenario is another important factor for effective evaluation. It must be built to reflect a realistic navigation

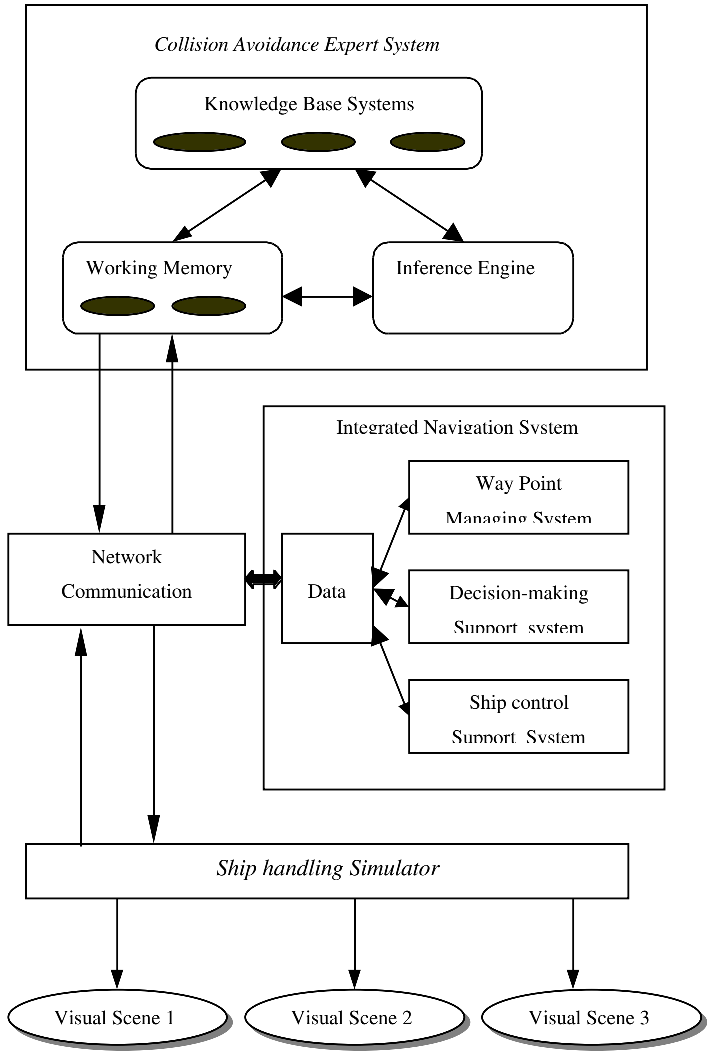

Fig.4 The Composition of Real Time Simulation System

environment. In this study, the simulation scenarios are designed as follows:

- Own ship particulars: Container: L=174m, B=26m, d=9.4m, Cb=0.57 and GRT = 16000GT;
- Navigating area: Coast area, congested Japan sea;
- Target ships are generated in terms of variable degree of difficulty for ship operation during the collision avoidance;
- Own ship is controlled by the developed expert system; and
- The own ship encounters 3 or 4 target ships on its course line frequently. These encounter situations are designed by navigators.

Using such simulation environment and scenarios, we asked navigators who have different navigating experiences to handle ships in two cases: manual ship handling and ship handling with the help of the developed expert system. Manual ship handling is carried out using visual information and radar information. Operators determine the timing of rudder operation, the action of collision avoidance, and so on. In the case of ship handling with the help of the developed expert system, operators only need to confirm the proposed action from the decision-making support system. During simulation, the system records all the ships' trajectories for later data analysis. In order to obtain the personal judgement and personal feeling about the system, we also interviewed the operators by a set of questionnaires on the effectiveness of decision-making support.

4.4 Simulation Result Analysis

After carrying out the simulation, we obtained simulation data and subjective judgement from domain experts. From these results, we derive the evaluating results for each evaluation item or index quantitatively or qualitatively. For fast-time traffic simulation, we focus on the ship trajectory analysis. Using simulated data, it is possible to show time history status such as the action of collision avoidance, action timing, safety distance between own ship and target ship, and so forth to domain experts. These results can help us to judge the correctness of collision avoidance action, the safety of ship navigation, and so on. For real-time simulation, we obtained the questionnaire results from operators about their subjective evaluation and simulation results. Operator‘s evaluation can be used to determine the qualitative indices such as the effectiveness of decision-making support information, the feasibility of proposed action and correctness of collision avoidance action and so forth. On quantitative analysis of simulation results, we focus on some main indices such as the safety level of own ship, the target ship risk corresponding to own ship action, and the deviation of course line. Due to space limitation, these results [9] can‘t included in this paper. Meanwhile, it is possible to compare the simulation result of manual operation with that using expert system support in order to prove the effectiveness of decision-making support during collision avoidance action.

Finally, in terms of the obtained evaluation indices, it is necessary to give an evaluating result for each evaluation item defined in the procedure. The following are our evaluating results:

- **Problem-solving ability for collision avoidance**

From simulation results of traffic simulation, the actions of the collision avoidance are reasonable and in accordance with the traffic regulations. In the light of the deviation of scheduled course line and target ship risk corresponding own ship's collision avoidance actions in the real-time ship handling simulation, the action of own ship does not cause any disturbance to target ships, and own ship can abide by traffic regulation to navigate in lane in the case of route navigation. Therefore, it can be said that the developed expert system possesses sophisticated problem-solving ability for collision avoidance.

- **Ship navigation safety**

From the results of the quantitative indices, it is found that the developed expert system can keep good safety level during navigation no matter of navigating area. In terms of the comparison of simulation results of manual ship handling and ship handling using the decision-making support system, we found that the safety level of own ship in the case of using the expert system is better than that in the case of manual ship handling.

- **Decision-making support ability**

According to the evaluation indices obtained from the simulation result analysis and operators' responses to questionnaire investigation, we can say that the developed expert system can provide effective support for operators' decision-making. As to the decision-making support information, the indication of the dangerous ship is very welcome and quite effective, but the information of the restricting ships and the indifferent ships is not useful for operator's decision-making. As a proposed maneuvering action for collision avoidance, the course line waypoints are effective and needed by operator.

Also, from the evaluation based on simulation, some problems such as flexibility of problem-solving, adaptability of high traffic density navigating area, etc. are found. According to the requirements of the system, it is necessary to improve the developed system.

5 Conclusions

In this paper, the authors proposed a simulation-based procedure for evaluating domain-oriented expert systems and discussed the evaluation of an expert system that we developed to assist ship navigators in their decision-making process to avoid collision. From the evaluation results obtained, it is concluded that the proposed evaluating procedure is effective for evaluating the developed expert systems. It reduces development cost and improves developing efficiency of the system. The proposed evaluating procedure can be applied to different domains such as aviation. Based on the authors' empirical experience, it is expected that criteria should be set up for evaluation indices. Therefore our future work is to study how to set up such criteria according to usability factors, evaluation environment, domain requirements, and system specifications.

References

1. Massimo Benerecett, Fausto Giunchiglia, and Luciano Serafini, *Multiagent System Verification via Model Checking*. Proceedings of EUROVAV 98, 1998
2. Catholijn M. Jonker, Jan Treur, and Wieke De Vries, *Compositional Verification of Agents in Dynamic Environment: a Case Study*. Proceedings of EUROVAV 98, 1998
3. Francky Trichet and Pierre Tchounikine, *Verifing and Validating a Task/Method Knowledge-base*. Proceedings of EUROVAV 98, 1998
4. Mysore Ramaswamy and Sumit Saeker, *Global Verification of Knowledge Based Systems via Local Verification of Partition*. Proceedings of EUROVAV 97, 1997
5. Frans Coene and Faul Dunne, *Verification and Validation of Rulebases Using a Binary Encoded Incidence Matrix Technique,* Proceedings of EUROVAV 97, 1997
6. Anca I. Vermesan, *Knowledge-based Systems: Verification and Validation in the view of Certification,* Proceedings of EUROVAV 97, 1997
7. Dieter Fensel and Arno Schonegge, *Specifying and Verifying Knowledge-based Systems with KIV,* Proceedings of EUROVAV 97, 1997
8. K. Kose and C. Yang, *A Collision Avoidance Expert System for Integrated Navigation System and Its Brush-up,* Journal of the Society of Naval architects of Japan. Vol. 177, 1995
9. C. Yang, *An Expert System for Collision Avoidance and Its Application,* Ph.D. Thesis, 1995
10. K. Kose, Y. Ishioka, and C. Yang, *A study on a Collision Avoidance Expert System for Decision-making Support in Integrated Navigation System,* Journal of The Society of Naval Architects of Japan. Vol.178, 1996
11. A. Nagasawa, K.Hara, K.Inoue, and K.Kose, The Subjective Difficulties of the Situation of Collision Avoidance II---- Toward the Rating by Simulation, Journal of Japan Institute of Navigation. Vol.91, 1993