Case Learning for CBR-Based Collision Avoidance Systems

Liu, Yuhong; Yang, Chunsheng; Yang, Yubin; Lin, Fuhua; Du, Xuanmin; Ito, Takayuki

Publisher's version / la version de l'éditeur:
The International Journal of Applied Intelligence, 2010-12-01

Case Learning for CBR-Based Collision Avoidance Systems

Yuhong Liu[1], Chunsheng Yang[2], Yubin Yang[3], Fuhua Lin[4], Xuanmin Du[5], and Takayuki Ito[6]

[1] Shanghai Aerospace Electronic Technology Institute, Shanghai, China

[2] Institute for Information Technology, National Research Council Canada, Canada

[3] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

[4] Athabasca University, Athabasca, AB, Canada,

[5] Shanghai Marine Electronic Equipment Research Institute, Shanghai, China

[6] Nagoya Institute of Technology, Nagoya, Japan

Abstract. With the rapid development of case-based reasoning (CBR) techniques, CBR has been widely applied to real-world applications such as collision avoidance systems. A successful CBR-based system relies on a high-quality case base, and a case creation technique for generating such a case base is highly required. In this paper, we propose an automated case learning method for CBR-based collision avoidance systems. Building on techniques from CBR and natural language processing, we developed a methodology for learning cases from maritime affair records. After giving an overview on the developed systems, we present the methodology and the experiments conducted in case creation and case evaluation. The experimental results demonstrated the usefulness and applicability of the case learning approach for generating cases from the historic maritime affair records.

Keywords: case-based reasoning, ship collision avoidance, maritime affair records, case learning, case base management.

1. Introduction

Human error is one of the most significant factors in maritime accidents. In particular, it was a root contributing factor for ship collisions in navigation. To avoid human error and improve navigation safety, many researchers [1-6] have focused on developing intelligent systems for collision avoidance. Yang et al. [4] developed a rule-based collision avoidance expert system based on the navigators' experiences and applied it to an integrated navigation system as a decision-making support system for collision avoidance. Similarly, Zhao and Wang [23] developed an intelligent decision support system for ship collision avoidance. These intelligent systems were developed based on rule-based reasoning techniques. The rules were created or obtained from traffic regulations, encounter scenarios, or navigation theories. However, such rules cannot fully mimic the human's ship-handling behaviors and experiences, which are the most important factor in ship-handling for collision avoidance. This is one of reasons that these research results are rarely applied to practical navigation systems.

To overcome the shortcomings of rule-based reasoning systems, we have started to look into applying Case-Based Reasoning (CBR) [28][29] and agent technologies to ship collision avoidance [15, 17, 18]. CBR is one of the reasoning paradigms and is a feasible and effective solution to problems which are difficult to be solved with traditional methods such as model-based reasoning. CBR-based approaches have been widely applied to various real-world applications such as diagnostics [26][32], design[27][34], planning [30], finance[31], health care [33], and decision-making support [7-12]. Also, we are looking into applying agent technologies to ship collision avoidance. A ship, navigating on an open and dynamic environment, can be treated as a rational and intelligent agent. Navigators on ships detect the changes of the environment, collect the information of other ships, judge the dangerous degree of the current situation, make decisions by using some knowledge, and take actions to avoid the collision with other ships or obstacles. To facilitate this research, we have developed a multi-agent system for ship collision avoidance. The agents in this system [17] were implemented with CBR-based decision making support for collision avoidance [18].

When we develop a CBR-based system for an application, a significant challenge that we face is case generation. Without a high-quality case base, it is impossible for a CBR-based system to work properly and effectively for solving the given problems. It is a challenge to automatically generate a high-quality case base because different applications require distinguished approaches for case generation. For example, Yang et al. [16] developed a methodology to automatically generate cases from the historic maintenance database for diagnostic CBR systems. To create the cases for the CBR-based collision avoidance systems, we looked into the usage of historic maritime affairs records which were collected over many years. These records documented either instructive and successful cases or edifying and failing cases. These cases are a valuable resource to generate cases for CBR-based collision systems. We collected many well-known maritime cases from Asia and Europe from 1976 to 2006. These case records were documented in an unstructured text format and in different languages, mainly in Chinese. To efficiently generate cases from these unstructured text records, we developed a methodology, focusing on Chinese text format, by using techniques from natural language processing (NLP) and CBR. In this paper, we present the developed techniques and some preliminary results.

The next section gives an overview on the CBR-based multi-agent system for collision avoidance; following that, we present a methodology developed for generating cases from maritime affair records; we then introduce the experiments along with some preliminary results following with a discussion section; the final section concludes the paper.

2.CBR-Based Collision Avoidance Systems

2.1 Multi-agent Systems for Collision Avoidance

In order to conduct the research for collision avoidance, we developed a multi-agent system [17] for simulating the real navigation environment. Figure 1 shows the system architecture. It consists of two types of agents: control agents and function agents. The control agents comprise system agents and union agents; and function agents are either ship agents or VTS (Vessel Traffic Service) agents. In general, the control agents manage the function agents, including information maintenance, agent communication, task partition and assignment, resource distribution and administration, conflict reconcilement, etc. A function agent performs CBR-based reasoning for collision avoidance by using the information from control agents and the environments. For example, when a potential collision risk arises, several related function agents are organized automatically to form a "dynamic society" that is a negotiation union, and work together to find a solution for the current problem. A union agent is assigned to each negotiation union. In some particular situation, a special function agent, VTS agent, is activated and licensed to take an action that must be obeyed by all the ship agents within the union.

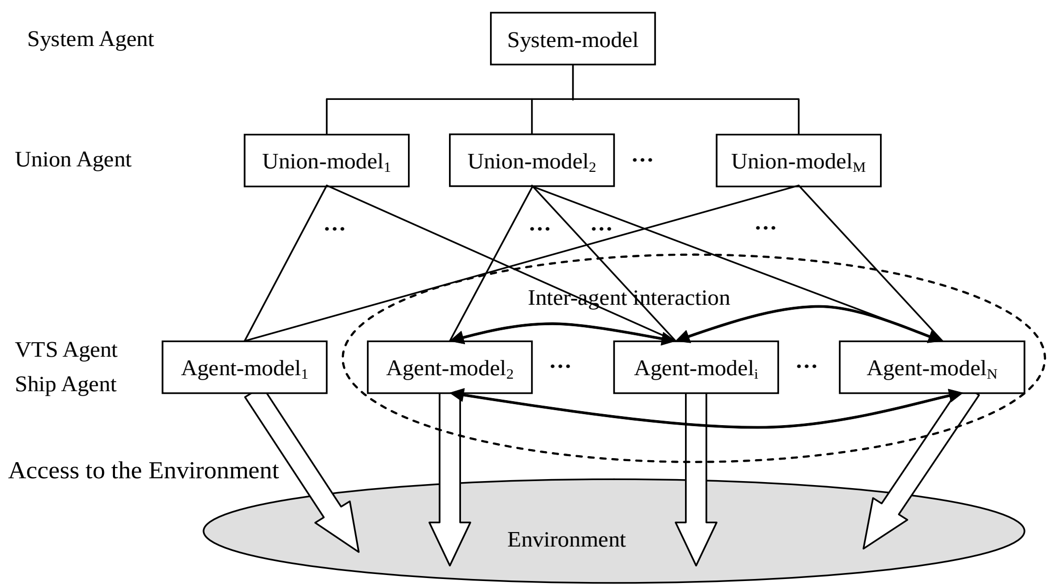


Figure 1: The architecture of multi-agent system for collision avoidance

To implement these agents, three different agent models are used: a system model, a union model, and an agent model. These models are designed following agent system technologies. Each agent model contains corresponding responsibility, state, data, and knowledge base. In particular, we implemented function agents with the CBR paradigm to provide decision-making support for collision avoidance. The details are presented in the following subsection. Communication that allows agents to interact each other for sharing information and achieving a common goal is implemented with ACL (Agent Communication Language), KQML (Knowledge Query Manipulation Language package [19], and AIS (Automatic Identification System) messages. The AIS is an onboard broadcast transponder system with which ships continually transmit their ID, position, course, speed over ground, ships static data and voyage related data to all other nearby ships and shore-side authorities on common VHF radio channels. AIS provides 22 messages to deliver the information, such as position reports, base station report, ship static and voyage related data, safety related message, and so on.

2.2 CBR-Based Collision Avoidance Systems

The function agents such as Ship agent and VTS agent in multi-agent systems shown in Figure 1 are implemented based on a BDI (Beliefs, Desires, and Intentions) model [25]. BDI agents are comprised of the following core data structures: Beliefs, Desires, Intentions, and plans. The BDI model has some philosophical basis in the Belief-Desire-Intention theory of human practical reasoning. We applied CBR to the BDI model for modeling human reasoning on collision avoidance. With the help of CBR, the architecture of CBR-based agents (Ship agent and VTS agent) is designed as shown in Figure 2. Basically, a CBR-based agent implemented in BDI model consists of two types of components: BDI function components and CBR function components. BDI function components such as communication, action trigger, and model base, provide capabilities to interact with other agents. CBR reasoning is performed through three main components: problem description, case retrieving, and case learning. The problem description component creates a collision problem based on real-time navigation data, including static information (ship type, ship length and sea gauge), dynamic information (navigating course, speed and position), and navigating information (the relative course and speed, azimuth, distance, DCPA, TCPA[1], encounter situation, and collision risk). Once a collision problem is defined, a case retrieval algorithm is used to retrieve similar cases from the case base which stores cases with a given presentation and an index structure. The case with a maximal value of similarity is selected as the proposed solution for the defined problem. Case learning is a key in CBR-based systems. The main task is to automatically generate cases from maritime affair records. We detail this core component in the following Sections.

[1] DCPA (Distance at the Closest Point of Approach) and TCPA (Time at the Closest Point of Approach) are important parameters for describing the related status between two ships in navigation. They are used as the criteria for deciding the action of avoiding collisions.

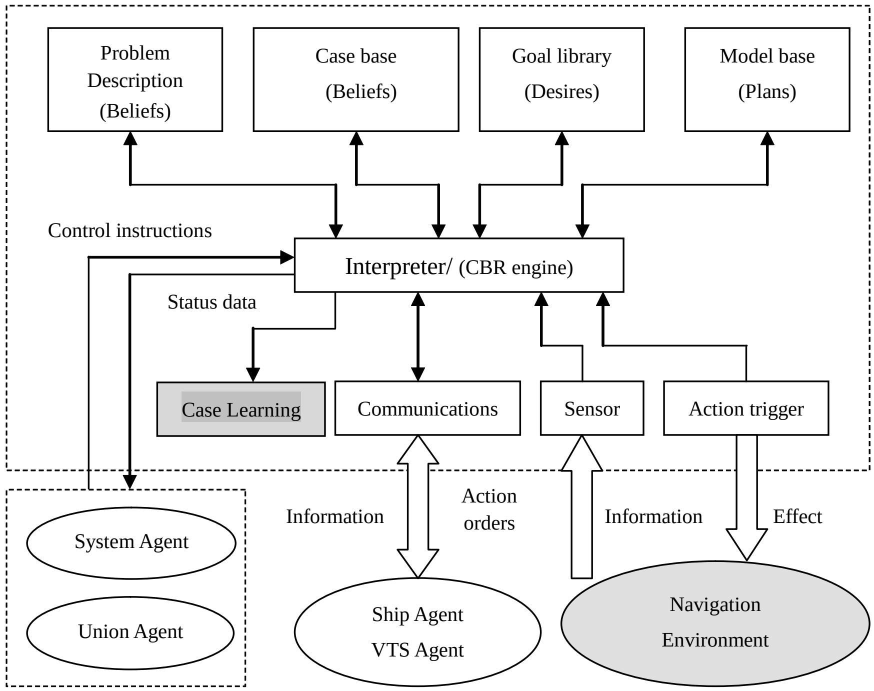


Figure 2: The paradigm of a CBR-based collision avoidance system

3.Automated Case Learning

Cases in CBR-based systems can be generated either in run time or at the initial stage of system development. For collision avoidance systems it is not feasible to generate cases in run time because of cost and safety issue. In our previous work [18], we proposed to create cases from ship-handling simulations. By analyzing the ship-handling trajectories obtained from a simulator we can create some cases for the given encountering situations. Other feasible and effective way is to generate cases from historic maritime affair records collected in navigation. This section presents a methodology developed to automatically generate cases from such records. We start from some preliminaries for a collision avoidance case based on practical navigation. Then we present the methodology.

3.1 Preliminaries

Ship collision avoidance is a dynamic process connecting with the sea, the ship, the human, and the environment, involving much information and many changes during a period. A ship collision avoidance case is a collection of ship operations over a long period. Figure 3 shows the dynamic encountering situation between Ship_agent1 and Ship_agent2. This dynamic course contains most information and changes related to two ships in navigation. The information is recorded at each view point for a given time. When we create cases from a given maritime affair record, we have to extract such information from the selected view point on the dynamic course.

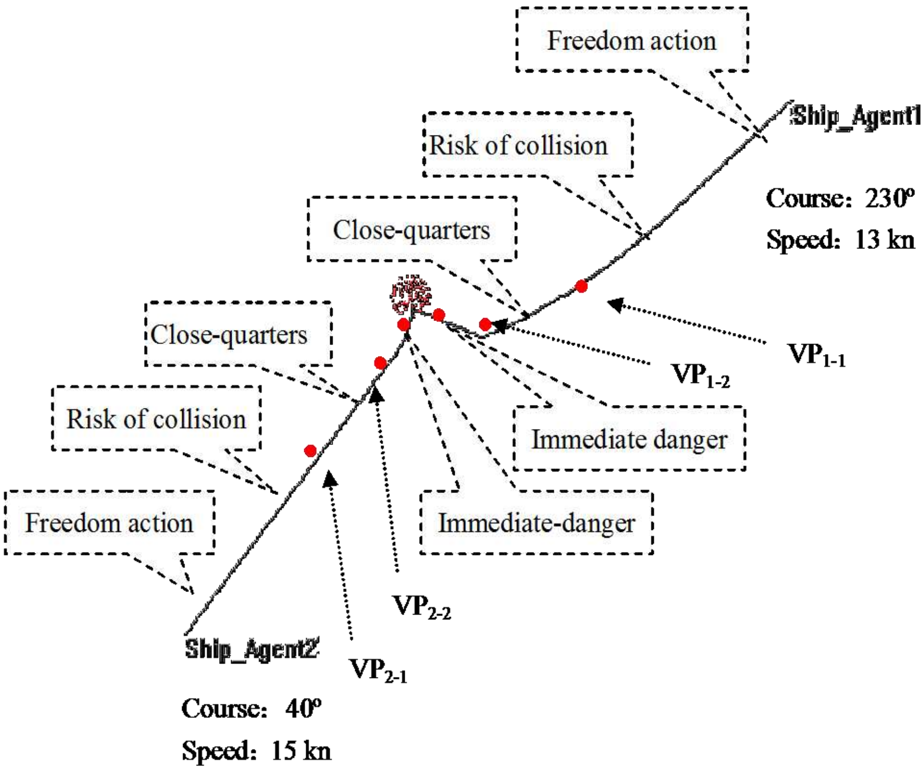


Figure 3: A dynamic course of a collision case

In order to describe the methodology, we first give some definitions on encountering situation, view point of navigation, maritime affair data, collision case, and case base.

Definition 1, **Encounter Scene (ES)**: a well-defined data structure. It is used to record the environment information (EI), the basic information (BI) of each ship, the relative information (RI) between own ship and each target ship and the proposed actions (PA) at a given time point. That is:

$$ES = < EI, BI, RI, PA > \quad (1)$$

Definition 2, **View Point (VP)**: during the ship collision avoidance, we label one of the encountering ships as the own ship (OS) and the others as target ships (TS). And then we select a time point (T) and record the encounter scene (ES) at this moment. VP is denoted as:

$$VP = < OS, TS, T, ES > \quad (2)$$

Definition 3, **Case base (CB)*:*** A case (c) is defined as $c = \{p_c, s_c, m_c\}$. p_c denotes a set of problem attributes, which describe a collision problem, and a set of VP; s_c is a set of solution attributes, either a single action or several actions for avoiding a collision; m_c contains all attributes related to case base maintenance, including redundancy, inconsistency, successful times, collision times, successful actions, and failed actions. Let CB denotes a case base, where $CB \supseteq \{c_1, c_2 \ldots\ldots, c_i, \ldots, c_n\}$.

Definition 4, **Maritime Affair Database (MD)*:*** For a given maritime affair record denoted as ***md**$_i$*, it contains the implicit or explicit environment description (ED) (sea state, weather condition, and visibility), the ship information description (SID) (encounter ships, ship name, type, length, draft, cargos and operation condition), and the collision or collision avoidance procedure description (PD) (the dynamic operation process and ship-ship relative information). For the collected maritime affair data, we denote it as a database, ***MD***, where $MD \supseteq \{md_1, md_2 \ldots\ldots, md_i, \ldots, md_q\}$.

3.2 Methodology

From the above definitions, our task is to create a CB given an MD. The developed methodology, which automates the procedures for case base creation, consists of three main processes:

- Identifying collision a problem and its solutions
- Creating a template case
- Updating the case base

A. Identifying a collision problem and its solutions

The task of this process is to find p_c and s_c given an ***md**$_i$* in ***MD.*** In this work, the collected ***MD*** is unstructured Chinese text. Such Chinese text format makes the work more complicated. Unlike English or other western languages, Chinese is character based, not word based. There are no “blank spaces” serving as word boundaries in Chinese sentences [20, 21]. In order to obtain ED, SID and PD information from a given ***md**$_i$*, we first conduct Chinese word segmentation, then perform semantic analysis based on a selected view point (VP).

Table 1, Algorithm for Automatic Segmentation

```
INPUT:          TextTree, DicBase;
OUTPUT:         SegBase;
INITIALIZATION:   NodeCount=0; DicCount=0; SegCount=0; SegFlg=FALSE; DelWord ←NULL;
                      RemainWord ←NULL;
-----------------------------------------------------------------------------
BEGIN:
  WHILE  (TextTree[NodeCount] is not NULL)    DO
  {                 InBuffer ← TextTree[NodeCount]; NodeCount++;
                IF InBuffer.Kind ∈{English, Number, Symbol and quotation }
                THEN {   SegBase[SegCount]← InBuffer; SegCount++; }
                 ELSE {   WHILE ( Infuffer.Words is not NULL) DO
                {   DicCount=0; SegFlg=FALSE;
                      WHILE (DicBase[DicCount] is not NULL]) DO
                         {   IF (InBuffer.Words == DicBase[DicCount])
                             THEN {   SegBase[SegCount] ← InBuffer.;
                                   SegCount++; SegFlg=TRUE;
                                   nBuffer ← DelWord;
                             } ELSE DicCount ++;   }
                           IF (SegFlg==FALSE)   THEN {
                                  FMaxMatch: DelWordPro (LastOne, InBuffer, RemainWord, DelWord);
                                  BMaxMatch: DelWordPro (FirstOne, InBuffer, RemainWord, DelWord);
                           IF (RemainWord is NULL) THEN
                           {   SegBase[SegCount] ← InBuffer.;
                               SegBase[SegCount].Kind= unknown;
                               SegCount++;
                            } ELSE InBuffer ← RemainWord; }
                         }   // end of   third WHILE
                }    // end of second WHILE
      }   // end of FIRST WHILE
   END
```

Chinese word segmentation separates a maritime affair record into 3 different paragraphs and extracts necessary information for ED, SID and PD from a given VP. The algorithm shown in Table 1 relies on a domain dictionary (*DicBase*) and a text tree (*TextTree*). The *DicBase* contains the main vocabulary for collision avoidance problem. It is created based on the following principles:

- The vocabulary is arranged in the sequence of WordKindSet (*WordKindSet* = <*noun, verb, adjective, adverb, conjunction, pronoun, preposition, auxiliary, quantifier, numeral*>);
- The words having same initial character are arranged in length sequence ordered from long to short;

- Each word in the vocabulary is associated with a numerical value to express its occurrence frequency.

A higher frequency is associated with a larger value. The *TextTree* is generated by paragraph segmentation, sentence segmentation and node segmentation. The three paragraphs corresponding to the ED, SID, and PD respectively are segmented. The sentence segmentation is based on the original text sentence and the segmentation tag is the five punctuations, “.”, “,”, “:”, “?” and “!”. The node segmentation is based on the kind of the character string. There are five kinds of characters, *NodeKindSet=<Chinese, English, number, special symbol, quotation mark>*. Only the Chinese characters need to be further segmented, and the other four kinds of character strings are treated as a solid semantic unit and need no more processing.

In Table 1, FMaxMatch and BMaxMatch are two simple algorithms for Chinese text match. They run in both forward and backward directions using the final word list as the references. Some domain knowledge is used in the algorithm to improve the segmentation efficiency. The outputs of FMaxMatch and BMaxMatch are stored in a static database, SegBase. The differences between the FMaxMatch and BMaxMatch outputs indicate the positions where the overlapping ambiguities occur. To reduce the ambiguity in segmentation, three rules are used to remove the ambiguity in the algorithm. The first rule is to remove overlapping ambiguity. The algorithm detects it and dispels it by selecting the words with higher occurrence frequency as the segmentation result or selecting the words manually. The second rule is to remove combination ambiguity by assigning a high priority to a combined string. The third rule is to deal with an unrecorded ambiguity string which is detected as *SegBase.Kind.* For an unrecorded ambiguity string, the “unknown” will be assigned to *SegBase.Kind.*

After performing the segmentation, we conduct semantic analysis based on the specific VP from the *SegBase.* The main purpose of semantic analysis is to determine the attribute values of p_c and s_c for a given key VP in the SegBase. A key VP contains important information or data as shown in Table 2. For some given VP, we may obtain the necessary information for p_c and s_c, some VP may not. Once we have p_c and s_c we start to create a case template as a potential case.

B. Creating a template case

Having p_c and s_c from the previous process, this process creates a potential case, $c_{tmp} = \{p_{ctmp}, s_{ctmp}, m_{ctmp}\}$ (where $p_{ctmp} \equiv p_c, s_{ctmp} \equiv s_c; m_{ctmp}$ is to be determined). A potential case is a structured case representation, which might be added to a case base as a new case or be merged with the other cases based on the case base maintenance policies. These policies are presented in following subsection

Table 2, An example of a VP selected for semantic analysis

Paragraph	Key point words		Value type
ED	Visibility		Chinese/numerical
	Wind power, course, speed		Chinese/numerical
	Flow course, speed		Chinese/numerical
	Snowfall, rainfall, Fog		Chinese/numerical
	Traffic density		Chinese
SID	Encounter ship number		Numerical
	Ship type		Chinese/English/mark
	Ship length, wide, sea gauge		Numerical
	Course, Speed		Numerical
PD	Ship to Ship	Distance, bear	Numerical
		DCPA/TCPA	Numerical
		Situation, stage	Chinese
		Collision risk	Chinese/numerical
	Operation	Steering	Chinese/numerical
		Speed change	Chinese/numerical
		Effect	Chinese

C. Updating case base

In ship navigation, an encounter situation (collision case) might occur several times, but the avoiding action may be either identical or different. In such a situation, we expect to create a single case to restore these experiences rather than multiple cases. Therefore, we need a sophisticated approach to manage the case base when we add a potential case to the existing case base. We use the algorithm shown in Table 3 for updating the created case bases.

Table 3, An algorithm for case base management

```
Input:  A given CB ⊇ {c_1, c_2 ...... c_i, ..., c_n}.
        A potential case c_tmp = {p_ctmp, s_ctmp, m_ctmp}
Process:
 For all c_i in CB   {
      If ¬∃ c_i similar to c_tmp        // where c_i = {p_ci, s_ci, m_ci}
           CB = CB ∪ c_tmp;             // add a new case to case base
      Else if (p_ci ≈= p_ctmp) ∩ (s_ci ≈= s_ctmp)
                 m_ci = upgradeMattributes( c_i );
      Else If (p_ci ≈= p_ctmp) ∩ (∀c_i ¬∃(s_ci ≈= s_ctmp))  do
              m_ci = upgradeMattributes( c_i );
              s_ci = s_ci ∪ s_tmp;
      End;
}
```

The goal of this algorithm is to determine the attributes of m_c for a given temporary case. The first step is to determine whether a potential case could be a new case. We check the redundancy or inconsistency of the potential case against the existing case base. If a case is not against any case in the existing case base, this case could be a new case, and will be added to the case base. Otherwise, we move on to the second step. It conducts case base management for the existing case base if we find a case (c_i) similar to c_{tmp}. Case base management includes updating an existing case, deleting a case, and merging multiple cases into a new case. This operation is realized by updating the attributes for m_c. If we detected a similar case (c_i) in the existing case base against the potential case c_{tmp}, i.e., $p_{ci} \approx= p_{ctmp}$ and $s_{ci} \approx= s_{ctmp}$, then m_{ci} will be updated to reflect the effect of the collision avoidance action applied to the collision problem. If c_{tmp} is a positive case, then we increase the count of successful actions of m_{ci} otherwise we increase the count of failed actions of m_{ci}. Similarly, if we detected a similar case (c_i) against case c_{tmp}, which has similar problem descriptions but different solutions, i.e., $p_{ci} \approx= p_{ctmp}$ and $s_{ci} \approx! s_{ctmp}$, the existing case will be updated by adding the new solution to it, so that the case will become more powerful for solving the similar collision problem in the future. In the algorithm, the similarity computation is supported by the developed CBR engine in our multi-agent system. The Appendix shows the formulas for computing the similarity between two cases.

4.Experiments and Case Evaluation

4.1 Experiments for case creation

We implemented a case learning system following the proposed methodology with the support of a CBR engine in our developed multi-agent system in a VC++ platform. Some experiments were conducted for creating cases from the collected maritime affair records.

We have collected almost 60 collision avoidance cases from maritime affair record books [22, 23, 24]. These records are written in Chinese and cover a time period from 1972 to 2006. Most of the records contain all information in unstructured Chinese text format. In other words, from the collision records we can extract necessary information for creating cases: ES (EI, BI, RI, PA), VP (OS, TS, T, ES), and actions taken for collision avoidance, or causes of collisions. Among 60 cases, 50 collision cases took place in Europe and were collected in Lloyd's Report, and 10 cases were from China. Most of collisions or collision avoidance situations took place near the coast and in shallow water areas. From the point of view of encounter ships, most of the cases are two ship collisions. Only five cases are related to multiple ship collisions. From the point of view of encounter relationships, 14 cases are heading collisions; 26 cases are crossing collisions; 10 cases are overtaking collisions; and 10 cases are out of navigating routes in shallow water areas. From the point of view of navigation environments, 27 cases happened under an invisible weather; and 33 cases under visible weather. We first created electronic versions for these records in Chinese text format. We then input these electronic documents to our developed case learning system and created cases automatically. In the end, we generated 58 cases successfully. Some cases are created from several collision cases because those cases may contain similar information on encountering situations and navigating environments, or took the similar action for avoiding the collisions. It is interesting that only 2 collision cases lack enough information for generating the cases.

4.2 Case evaluation

Although the developed methodology can learn the valuable cases from historic maritime affair records, the created cases have to be evaluated carefully before they are deployed to CBR-based collision avoidance systems. To evaluate the cases, we have developed an evaluation system, which is also incorporated into the multi-agent systems. This system is capable of evaluating the ship-handling results by analyzing the trajectories collected during ship-handling. Figure 4 shows the developed system [14]. Basically, the system consists of three components to perform three main tasks: data processing, risk analyzing, and knowledge reasoning. Here is a brief description of each task.

Data processing extracts the necessary data from ship trajectories and computes some derived parameters from original data. Original data from ship trajectories include the longitude, latitude, course and speed of own ship and target ships, as well as the data related to winds, currents and navigating areas at the given time. The derived parameters include the relative information such as the bearing and distance between two ships. These parameters help evaluate the risk of ship-handling actions.

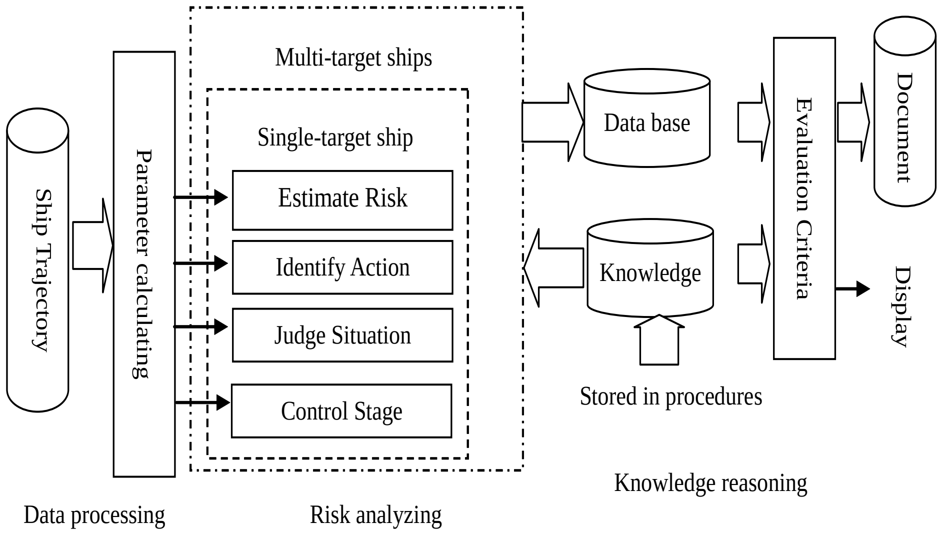


Figure 4: Evaluation system for ship trajectory

Risk analyzing estimates the risk of ship at a given time on the trajectory. In this system, four kinds of evaluation criteria are developed. They are action identification, action timing measurement (stage control), encountering situation judgment, and risk estimation. The former three criteria are determined or measured using DCPA, TCPA, and navigation regulations. The risk is computed using Equation 3 and 4. Equation 3 is used to compute the risk between own ship and one single target ship. Equation 4 computes the risk for multi-target ships.

$$\mathrm{Risk(i)} = \begin{cases} 0 & \mathrm{DCPA_i} \geq \lambda_{\mathrm{DCPA}} \\ \dfrac{1 - e^{2\cdot(\mathrm{DCPA_i} - \lambda_{\mathrm{DCPA}})^2}}{2} + \dfrac{e^{2\cdot \mathrm{TCPA_i}^2}}{2} & \mathrm{DCPA_i} < \lambda_{\mathrm{DCPA}} \end{cases} \quad \text{….. (3)}$$

$$\mathrm{Risk} = \frac{1}{N} \cdot \sum_{i=1}^{N} \mathrm{Risk(i)} \quad \text{…. (4)}$$

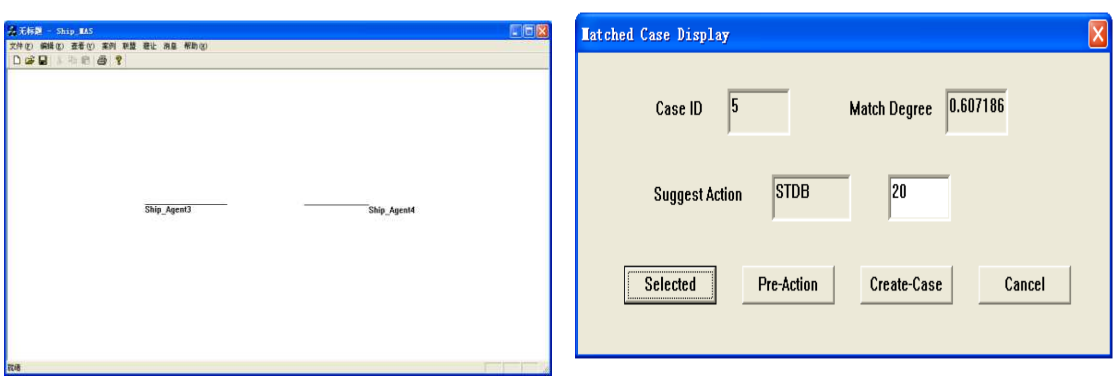
Where:

- λ_{DCPA} is the threshold of *DCPA* and has different values under the condition of in sight and restricted visibility
- N is the number of target ships

Knowledge reasoning makes the final judgment for the ship trajectory by using the risk estimation, a grade criterion and some rules. Usually this will give a grade as a final result for a given ship-handling simulation. This function is mainly used to evaluate the ship-handling simulation in training students. We do not use it for case evaluation.

In this study, the objective is to evaluate the usefulness of the proposed methods for learning cases from historic maritime affair records. We are interested in the risk estimation for given trajectories. Therefore, we first conducted ship-handling simulations with the created cases. To simplify the procedures, we only consider a single target ship encounter situation for our evaluation task. While running ship-handling simulations using the designed encounter scenario, we collected ship trajectories for each ship. In other words, the cases created from the experiments are stored in the case bases in the developed CBR-based agent systems. Then we designed some encounter scenarios for ship-handling simulation systems which may run into the collision situation. Based on the given problem descriptions, the system retrieves the similar case for avoiding collision. During the simulation, ship-handling trajectories of a target ship and own ship are collected for analysis. Figure 5 shows an example of case retrieving in the case evaluation. In this experiment, the system retrieved a similar case from the case base given collision problem. Figure 5(*a*) shows an encounter situation, in which two ships are head on to each other. They are noted as Ship_Agent3 and Ship_Agent4. Ship_Agent3 has course 90º and speed 12 kn. Ship_Agent4 has course 270º and speed 10kn. In order to formula a problem representation for this collision situation, Ship_Agent3 is defined as an own ship. Figure 5(*b*) shows that a case is retrieved from the case base in terms of the computed similarity. As a result, the case with ID 5 is retrieved since it has the highest value of similarity, 0.607186. Hence, its solution (turn right 20º) from the retrieved case is used as a proposed solution for the current collision problem. The ship-handling system applies the solution to handle ship for collision avoidance. Finally the risk is computed by Equation 3 to judge if the collision avoidance action is safe and effective. In this situation, the risk was computed as zero and the quality of the case is considered as "good case".

For 58 cases generated in case creation experiments, we evaluated them one by one as we described above. Table 4 shows the evaluation results for 58 cases. As we mentioned, we only analyze the risk for each ship-handling simulation trajectory. We did not take the visibility into risk analysis. The evaluation results are shown in Table 4, in which the number in each column represents the "good" case for the given criteria. Not surprisingly, 56 of 58 cases can provide a



(a) **Figure 5: An example of case evaluation** (b) safe action for collision avoidance. Only two cases from “overtaking” and “out of route” fail to provide a right action or right timing for collision avoidance.

Table 4, The preliminary results for 58 cases evaluation

58 CASES CREATED FROM EXPERIMENTS		RISK ANALYSIS IN CASE EVALUATION			
		Risk estimation	Action taken	Action timing	Stage Control
Encounter Situation	Heading	14	14	14	14
	Crossing	24	24	24	24
	Overtaking	9	9	9	9
	Out of route	9	9	9	9
Total “good” cases		56	56	56	56

5.Discussion

From the case creation and evaluation experiments, it is obvious that the cases were successfully created from maritime affair records using the proposed methods. The preliminary results demonstrated the usefulness and feasibility of the developed techniques.

It is worth noting that the evaluation method in this work is very simple. Only risk analysis is performed to judge the quality of cases. While computing the risk, we only take TCPA and DCPA into account. This is limited because the safety of ship navigation involves many factors such as human experience, navigating environment, and evaluation criteria. For example, the better risk estimation should involve the impact of visibility. In other words, λ_{DCPA} should be determined based on the visibility of navigating environment. It is a complicated procedure [4]. We argue that more practical ship-handling evaluations are necessary, and the experienced navigators should be involved in order to fully evaluate the quality of cases.

It is also worth noting that the developed techniques are easily transferable from Chinese language text processing to other language text processing by applying the corresponding domain dictionary and text tree, preprocessing and morphological analysis, grammar and parsing , and semantic interpretation for given language such as English.

In this work, we used only a small set of maritime affair records to validate the usefulness of the developed techniques. Several issues remain open, including safety criteria for case evaluation, more practical methods for case evaluation, and case uncertainty in retrieving case for decision-making on collision avoidance. All these issues will be our future work.

A CBR-based approach for collision avoidance is a problem-solving paradigm, which solves collision problem based on the past experience stored in the case base. It is not a modeling of the collision problem. Therefore, there is a limitation, i.e., for a new collision situation, the system may not be able to find a right solution if the case base does not provide the similar collision case in the past. Considering such a limitation, the developed techniques should be used as a decision-making support tool, instead of an automated collision avoidance system, in the real world navigation.

6.Conclusion

In this paper, we started from an overview on the developed CBR-based agent systems for collision avoidance. We introduced a methodology proposed for automatic case generation, which was developed using techniques from Chinese Language Processing and CBR. Even though the methodology relies on Chinese language processing techniques, it is easy to transfer to English or other languages. We conducted the experiments for learning cases from the collected 60 maritime affair records. For the created cases, we performed the evaluation by using the developed evaluation system and ship-handling simulator. The preliminary results show that the proposed method is useful and effective for creating cases from the historic maritime affair records.

Acknowledgments

Part of work was done while Dr. Yuhong Liu visited the National Research Council Canada in 2007. This work is supported in part by the National Natural Science Foundation of P. R. China (Grant Nos. 60875011, 60723003, 60721002), and the Natural Science Foundation of Jiangsu Province, P. R. China (Grant No. BK2007520).

Appendix A: Similarity Computing

If $V_{ck}^{i} \in [0,1]$ is defined as the similarity between a current problem *c* and a stored case *k* in the *i*th attribute, x_c^i and x_k^i are the value for the i^{th} attribute in *c* and *k* respectively. With different value types, PV (Precise Value), FV (Fuzzy Value) or DV (Default Value), there are four kinds of functions to calculate attribute similarity.

1) Similarity between two precise values: for two precise values x_c^i and x_k^i, the equal or unequal judgement method is not adopted. Instead, a more flexible similarity function is introduced.

$$V_{ck}^{i} = 1.0 - \frac{\left|x_c^i - x_k^i\right|}{\lambda_i} \tag{5}$$

where: λ_i is a threshold for the i^{th} attribute.

2) Similarity between a precise value and a fuzzy value: for two values x_c^i and x_k^i, if one is a precise value *x*, and the other is a fuzzy value with a fuzzy set U, then the fuzzy membership function $\mu_U(x)$ is selected as the similarity function.

$$V_{ck}^{i} = \mu_U(x) \tag{6}$$

3) Similarity between two fuzzy values: for two fuzzy values x_c^i and x_k^i with two fuzzy set *A* and *B*, their approximate relation matrix can be calculated through equation (5) and (6).

$$V_{ck}^{i} = \rho \cdot \left[A \circ B + \left(1 - A \otimes B\right)\right] \tag{7}$$

where:

$$\begin{cases} A \circ B = \underset{x \in X}{\vee}\left(A(x) \wedge B(x)\right) \\ A \otimes B = \underset{x \in X}{\wedge}\left(A(x) \vee B(x)\right) \\ \rho = \frac{1}{2}\left(1 - \left|CG_A - CG_B\right|\right) \end{cases} \tag{8}$$

CG_A and CG_B are the centre of gravity of *A* and *B*.

4) Similarity between default value and non-default value: for two values x_c^i and x_k^i, if one is the default value NULL, and the other is a non-default value, then similarity function will be the default value because default value NULL can be any value in the algorithm:

$$V_{ck}^{i} = 1.0 \tag{9}$$

Finally, S_{ck} , the similarity between a current problem *c* and a stored case *k*, can be obtained by equation (8)

$$S_{ck} = \frac{\sum_{i=1}^{m} \omega_i \cdot V_{ck}^{i}}{\sum_{i=1}^{m} \omega_i} \tag{10}$$

where: $\omega_i \in [0,1]$ is the weight of the i^{th} attribute and *m* is a total of attributes in a given case.

References

[1] Y. Sato and H. Ishii, "Study of Collision-avoidance System for Ships", Control Engineering Practice, Vol. 6, pp. 1141-1149, 1998.

[2] Y. Liu and W. Yang, "The Structure Design of an Intelligent Decision Support System for Navigation Collision Avoidance", IEEE Third International Conference on Machine Learning and Cybernetics, Vol. 1, pp. 302-307, August 2004.

[3] Y. Liu, "A Design and Study on Intelligence Collision Prevention Expert System for Navigation", *Ph.D. Thesis, Harbin Engineering University*, *China*, July 1999.

[4] C. Yang, "An Expert System for Collision Avoidance and Its Application", *Ph.D. Thesis, Hiroshima University, Japan,* September 1995.

[5] C. H. Wang, "The Integrated Design of Fuzzy Collision Avoidance and H_∞- Autopilots on ships", *The Journal of navigation*, Vol. 55, No.1, pp.117-136, 2002.

[6] Y. Liu, X. Du and S. Yang, "The Design of a Fuzzy-Neural Network for Ship Collision Avoidance", *ICMLC 2005*, *Lecture Notes in Computer Science,* Vol. 3930, pp. 804-812, July 2006.

[7] I. Watson and F. Marir, "Case-Base Reasoning: A Review", Knowledge Engineering Review, Vol. 9, No. 4, 1994

[8] C. Yang, B. Orchard, B. Farley and M. Zaluski, "Automated Case Base Creation and Management", The Proceedings of International Conference on Industrial & Engineering. Applications of Artificial Intelligence & Expert System (IEA/AIE-2003), June 2003.

[9] M, Therani, J. Zhao and B. Marshall, "A Case-based Reasoning Framework for Workflow Model Management", Data & Knowledge Engineering, Vol. 50, pp. 87-115, 2004.

[10] P. Avesani, S. Ferrari and A. Susi, "Case-Based Ranking for Decision Support Systems", ICCBR 2003, LNAI 2689, pp. 35-49, 2003.

[11] D. McSherry, "Similarity and Compromise", ICCBR 2003, LNAI 2689, pp. 291-305, 2003.

[12] J. Nordlund and H. Schafer, "Case-Based Reasoning in a Support System", Master's Thesis, UMEA University, Sweden, April 2006.

[13] Y. Liu and S. Hu, "An Evaluation System for Single-Target Ship Collision Avoidance Based on Data Fusion", NAVIGATION OF CHINA, Vol. 65, No. 4, pp. 40-45, Dec. 2005.

[14] Y. Liu and H. Liu, "Case Learning Based on Evaluation System for Vessel Collision Avoidance", The Proceedings of IEEE The Fifth International Conference on Machine Learning and Cybernetics, Vol. 4, pp. 2064-2069, August 2006

[15] Y. Liu, M. Wen and Z. Du, "A Case Learning Model for Ship Collision Avoidance Based on Automatic Text Analysis", The Proceedings of IEEE The Fifth International Conference on Machine Learning and Cybernetics, July, 2009, Baoding, China.

[16] C. Yang, B. Farley and B. Orchard, "Automated Case Creation and Management for Diagnostic CBR Systems", International Journal of Applied intelligence, Vol. 28, No. 1, 2008, pp. 17-28

[17] Y. Liu, C. Yang and X. Du," Multi-agent Planning for Ship Collision Avoidance", The Proceedings of IEEE International Conferences on Cybernetics & Intelligent Systems (CIS) and Robotics, Automation & Mechatronics (RAM) (CIS-RAM 2008), Chengdu, China, June, 2008.

[18] Y. Liu, C. Yang and X. Du, "A CBR-based Approach for Ship Collision Avoidance ", The Proceedings of the 21st International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems (IEA/AIE-2008), Wroclaw, Poland, June 2008

[19] Y. Labrou, "Standardizing Agent Communication. Multi-agents Systems and Applications", Lecture Notes in Computer Science. (2001) pp.74～9

[20] H. Gong, C. Gong and C. Zhou, "Chinese Word Segmentation System Research", The Journal of Beijing Institute of Machinery, Vol. 19, No. 3, pp.52-61, Sep. 2004.

[21] J. Qiu, T. Wen, and L. Zhou, "Research of Chinese Automatic Segmentation and Content Analysis Method", The Journal of The China Society For Scientific and Technical Information, Vol. 24, No. 3, pp.309-317, June. 2005

[22] China MSA (Maritime Safety Administration), "Typical Water Traffic Accident Cases", China Communications Press, pp.1-89, Oct. 2007.

[23] J. Zhao and F. Wang, "The Collision Regulations and Cases", Dalian Maritime University Press, pp.560-685, Mar.1997.

[24] E. J. Buzek and H. M. C. Holder, "Collision Cases Judgments and Diagrams", Lloyd's of London Press Ltd. 1984.

[25] A. S. Rao and M. P. Georgeff, "Modeling Rational Agents within a BDI-Architecture", *The proceedings of The 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pp. 473-484, 1991.

[26] M. Park, K. K. Lee, K. M. Shon, and W. C. Yoon, "Automating the Diagnosis and Rectification of Deflection Yoke Production Using Hybrid Knowledge Acquisition and Case-Based Reasoning", International Journal of Applied intelligence, Vol. 15, No. 1, 2001, pp. 25-40

[27] D. B. Leake and D. C. Wilson, "A Case-Based Framework for Interactive Capture and Reuse of Design Knowledge", International Journal of Applied intelligence, Vol. 14, No. 1, 2001, pp. 77-94

[28] S. C. K. Shiu and S. K. Pal, "Case-Based Reasoning: Concepts, Features and Soft Computing", International Journal of Applied intelligence, Vol. 21, No. 3, 2004, pp. 233-238

[29] C.A. Policastro, A. C. P. L. F. Carvalho, and A. C. B. Delbem, "A Hybrid Case Adaptation Approach for Case-based Reasoning", International Journal of Applied intelligence, Vol. 28, No. 2, 2008, pp. 101-119

[30] P. Avesani, A. Perini, and F. Ricci, "Interactive Case-Based Planning for Forest Fire Management", International Journal of Applied intelligence, Vol. 13, No. 1, 2000, pp. 41-57

[31] K. Kim, “Toward Global Optimization of Case-Based Reasoning Systems for Financial Forecasting”, International Journal of Applied intelligence, Vol. 21, No. 3, 2000, pp. 239-249

[32] M.P. Féret and J.I. Glasgow, “Combining Case-Based and Model-Based Reasoning for the Diagnosis of Complex Devices”, International Journal of Applied intelligence, Vol. 7, No. 1, 1997, pp. 57-78

[33] I. Bichindaritz, S. Montani, and L. Portinale, “Special Issue on Case-based Reasoning in the Health Sciences”, International Journal of Applied intelligence, Vol. 28, No. 3, 2007

[34] S. Kravis and R. Irrgang, “A Case Based System for Oil and Gas Well Design with Risk Assessment”, International Journal of Applied intelligence, Vol. 23, No. 1, 2005, pp.39-53