Scientific Journals
Maritime University of Szczecin

Zeszyty Naukowe
Akademia Morska w Szczecinie

2012, 30(102) pp. 18–24

2012, 30(102) s. 18–24

Methods of knowledge representation contained in COLREGS

Metody reprezentacji wiedzy zawartej w MPDM

Marcin Breitsprecher

Maritime University of Szczecin, Faculty of Navigation, Institute of Marine Technologies
Akademia Morska w Szczecinie, Wydział Nawigacyjny, Instytut Technologii Morskich
70-500 Szczecin, ul. Wały Chrobrego 1–2, e-mail: bremar@am.szczecin.pl

Key words: expert system, knowledge base, knowledge representation

Abstract
The article presents the problem of knowledge representation contained in COLREGS for its use in navigation information systems. The idea of knowledge and representation of knowledge was discussed. A comparison of selected methods of knowledge representation was made and possibility of their use to represent the COLREGS rules was considered. By using selected COLREGS rule and presented methods of knowledge representation, functionality of the knowledge base was analyzed.

Słowa kluczowe: system ekspertowy, baza wiedzy, reprezentacja wiedzy

Abstrakt
W artykule przedstawiono zagadnienie reprezentacji wiedzy zawartej w MPDM dla potrzeb jej wykorzystania w nawigacyjnych systemach informacyjnych. Omówiono pojęcie wiedzy oraz sposoby jej reprezentowania. Dokonano porównania wybranych metod reprezentacji wiedzy. Rozpatrzono możliwość ich zastosowania do reprezentacji prawideł MPDM. Na przykładzie wybranego prawidła przeanalizowano funkcjonalność bazy wiedzy utworzonej z użyciem przedstawionych metod reprezentacji wiedzy. Sformułowano wnioski.

Introduction

Knowledge is a broad concept which is defined in many ways in the context of the chosen discipline. Encyclopedic definition in the broad sense defines knowledge as "any collection of information, views, beliefs, with cognitive and/or practical value", while the narrower terms as "generally reliable information about the reality and the ability to use them" [1]. Knowledge is treated as a collection of information from a particular area, which was based on experience and in the learning process. Knowledge is a symbolic description of the surrounding world and the phenomena occurring around [2]. One of the types of knowledge is an expert knowledge. This is the knowledge of specified discipline. Having the expert knowledge enables noting patterns and structures specific to a particular problem [3].

The main task of navigational information systems is acquisition and presentation of navigation information in order to aid the navigator in the decision making process. Implementing the knowledge of safe operation of the ship may help increase functionality of such systems for decision support. This includes the interpretation of COLREGS rules. Implementation of COLREGS in decision support system will allow interpreting the situation and determining navigational manoeuvres in collision situations, taking into account existing regulations. This also applies to local regulations in certain areas (such as fairways, ports). This would minimize the most common cause of accidents – human error [4].

Acquiring and implementing expert knowledge in the field of navigation in navigation systems is essential to develop more effective tools for navigator, which is associated with a desire to expand the navigation information systems into navigational decision support systems. This requires the formalization of knowledge and then building a knowledge base.

Creating a knowledge base is a multi-step task in which the following phases can be distinguished [2]:

- identification – identify the problem to solve;
- representation – analysis of methods for knowledge representation;
- formalization – creation of knowledge structures;
- implementation – a combination of formal knowledge and inference engine;
- testing – validation of the knowledge base.

The acquisition, processing and using of knowledge is covered by knowledge engineering [5].

An important source of information for navigators is COLREGS regulations. Navigational Decision Support System developed at the Maritime Academy in Szczecin is an example of the use of navigational knowledge in this area [6, 7]. This system implements the above mentioned provisions. There are still studies to develop new methods and ways of navigational knowledge representation. This would improve the process of knowledge modification and verification, including the possibility of supplementing it with local regulations.

The formulation of the problem

COLREGS regulations provide a set of laws governing the navigators in various navigational situations. These regulations are difficult to implement into the system directly. The acquisition and representation of knowledge in this field requires an interpretation of the expert, to identify premises and conclusions or defining and verifying the rule set.

As an example Rule No. 13 was considered, which deals with the situation of overtaking. Based on an analysis of records of that regulation it can be concluded that [8]:

1) any vessel overtaking any other shall keep out of the way of the vessel being overtaken;
2) statement No. 1 does not apply when the overtaking vessel is not under command;
3) the overtaking vessel is one that approaches the overtaken vessel from a direction more than 22.5 degrees abaft her beam;
4) if the navigator is in doubt whether he overtakes another vessel he should assume so and proceed accordingly;
5) any changes of the bearing between the two vessels shall not alter the initial state, e.g. to make the overtaking vessel a crossing vessel;
6) the lateral distance between ships is essential in case of overtaking on parallel courses.

Above considerations show that regulation interpretation process should consider:

– bearing of the overtaken vessel;
– bearing of the overtaking vessel;
– navigational status;
– speed of vessels;
– lateral distance between the vessels.

Implementing the knowledge contained in the analyzed rule requires the use of appropriate methods of knowledge representation. This also applies to all other rules.

Selecting a particular method of knowledge representation determines the effectiveness of using acquired knowledge. It also determines, among others how to implement the remaining elements of the system, such as inference engine.

Selected methods of knowledge representation

Knowledge representation methods include the topics of modeling real-world by using computer systems. The most commonly used techniques for organizing knowledge are [2]:

– based on the direct application of logic: propositional calculus, predicate calculus;
– writing statements using semantic networks and frames;
– decision rules (including the vectors of knowledge);
– decision trees;
– using computational models.

Knowledge representation methods are aimed in particular [9]:

1) to prepare knowledge in appropriate format, which allows the use of a computer system;
2) to store and maintain knowledge in a form as close to the knowledge given by an expert;
3) to present knowledge in such a way that they can be modified (addition of rules and facts).

Decision trees. The decision tree is a graphical method for decision support. By definition, a tree is composed of the root, nodes, branches (possible variants) and leaves. The attributes are stored in nodes, the branches represent the values of these attributes, while the leaves represent different classes of decision. Specific modifications of decision trees are decision diagrams. In case of diagrams access to a specific node is possible by using more than one path.

The decision tree is constructed by using a training set, which contains objects (navigational situations). These objects are described with attributes and assigned to particular decision classes. In order to construct the tree, it is necessary to examine the information contained in the training set. Calculations cover probability of each decision class, the entropy of the system (the entire training set) and significance (entropy) of the individual attributes.

Entropy of the training set is expressed by the formula:

$$I(s_1, s_2, \ldots, s_m) = -\sum_{i=1}^{m} p_i \log_2(p_i) \tag{1}$$

where:

s_i – number of objects in classes;

m – number of classes, where objects are classified;

p_i – probability that the selected object belongs to the class C_i and is $0 < p_i < 1$.

The entropy of attribute A in terms of decision class division $\{C_1,...C_m\}$ was determined using the formula:

$$E(A) = \sum_{k=1}^{m} \frac{(s_{1k} + s_{2k} + \ldots + s_{mk})}{s} I_{A,k}(s_{1k}, s_{2k}, \ldots, s_{mk}) \tag{2}$$

The smallest entropy value calculated for each attribute shows which one is most important. The construction of decision tree then starts from this attribute.

Rule sets representation of knowledge. The most common method of knowledge representation in expert systems are decision rules [2]. It should be noted that a set of statements or facts are not sufficient to describe any field of knowledge. From some certain facts other facts may be inferred. Then those relations might be written as a set of rules. The knowledge base is a set of rules and a set of facts. A rule consists of premises and conclusion. The expressions in the premisses and conclusions are called clauses.

The general form of rules:

$$\text{IF } X \text{ is } F \text{ THEN } Y \text{ is } G \tag{3}$$

If the premises or the conclusions consist of more than one argument, they can contain logical operators AND, OR and NOT, for example:

$$\begin{gathered}\text{IF } X1 \text{ is } F1 \text{ AND } X2 \text{ IS } F2 \\ \text{OR} \ldots Xn \text{ IS } Fn \text{ THEN } Y \text{ is } G\end{gathered} \tag{4}$$

In more formal notation, it drops the words IF–THEN and the implications of using the symbol is written, respectively:

$$\begin{gathered}(X, F) \Rightarrow (Y, G) \text{ and for more arguments} \\ (X1, F1) \wedge (X2, F2) \vee \ldots (Xn, Fn) \Rightarrow (Y, G).\end{gathered} \tag{5}$$

Machine learning algorithms. See5/C5.0 algorithm is derived directly from the developed in 1993 C4.5 algorithm by Ross Quinlan [10]. The algorithm processes data set prepared in the form of a set of learning examples. Training set is created in the representation of attribute-value objects and presents it in the form of vectors of attribute-value--decision class. The algorithm uses so-called data mining mechanism, and as a result gives a classifier in form of decision tree or set of rules. The general algorithm pseudocode is as follows [11]:

1. Check the set of cases;
2. For each attribute *a*;
   1. Find the information gain after splitting the set on attribute *a*;
3. Let *a_best* be the attribute with maximum normalized information;
4. Create a decision *node* using *a_best*;
5. Recurse on the subset obtained by splitting set on *a_best* and add nodes as descendants of the *node*.

Selected methods of knowledge representation for navigational knowledge

Methods of knowledge representation presented in chapter 2 were used to represent the knowledge contained in Rule 13 COLREGS. The methods were analyzed in order to compare possibility of implementation and their effectiveness for completion and verification of created knowledge base.

Decison tree

To create a decision tree training set was prepared. The set describes the navigational situations during the meeting of two vessels. Cases when own ship approaches the target ship from a direction abaft her beam are considered. The question is whether own ship is “give way” or “stand on” vessel. The answer is given on basis of attributes contained in the training set and decision class assigned to each object.

Training set consists of objects presented in table 1. Particular attributes are described below:

speed – own ship’s speed in relation to the target ship (equal, less, more);

range_>_5NM – lateral distance between ships (false – smaller than 5 NM, true – greater than 5 NM);

course_aspect – dependence of courses between vessels (divergent, intersect, parallel). The value of the attribute is specified explicitly;

status_dmg – the status “not under command” of own ship (false, true).

Objects were classified into two decision classes: C1-you are “give way” vessel and C2-you are “stand on vessel”.

Table 1. The data for decision tree learning – fragment
Tabela 1. Dane uczące dla drzewa decyzyjnego – fragment

Object	Speed	range_> _5NM	course _aspect	status _dmg	Class (decision)
1	equal	false	divergent	false	C2-you are stand on vessel
2	equal	false	parallel	false	C2-you are stand on vessel
3	less	false	parallel	false	C2-you are stand on vessel
4	less	true	parallel	false	C2-you are stand on vessel
...	more	false	divergent	false	C2-you are stand on vessel
10	more	false	intersect	false	C1-you are give way vessel
11	more	false	intersect	true	C2-you are stand on vessel
12	more	false	parallel	false	C1-you are give way vessel

Probabilities, entropy of the system and the importance of attributes were calculated by using object from the training set. Values of p_1 and p_2 determine the probability of the objects belonging to one of the decision class C1, or C2. The results of calculations based on formulas (1) and (2) are as follows:

$p_1 = 0.20$
$p_2 = 0.80$
$I(s_1,s_2) = 0.7219$
$E_{(\text{speed})} = 0.4598$
$E_{(\text{range\_>\_5NM})} = 0.7185$
$E_{(\text{course\_aspect})} = 0.5837$
$E_{(\text{status\_dmg})} = 0.6996$

Calculation of the entropy shows that the most important attribute, that contains the greatest amount of information to classify objects into decision classes is *speed*. This means that the construction of decision tree should start with this attribute.

Based on the results decision tree was constructed as shown in figure 1.

It should be noted that the decision tree for much more complicated problems can be constructed in many variants. For example, a tree structure containing all the provisions COLREGS would be considerably more complicated. This will lead to expand the tree structure and increase the requirements for its implementation. Another problem is the issue of adding new knowledge to decision tree. This process usually results in the need to rebuild the tree, or build it from scratch.

Decision Rules

Decision rules can be derived directly from decision trees. Such set of rules are obtained by writing down the attributes in nodes and values from branches, descending from the root to the leaves which represent decision classes.

For example, based on previously constructed decision tree, the following set of rules was created. Then, it was stored in the knowledge base of rule-model RMSE expert system [12]. The rule in this system has the following form:

rule (consecutive_number, „*conclusion*”, [„*premise*”], display_semaphore)

and can be understood as follows: If the *premise* is satisfied, then the *conclusion* is true.

Created set of rules is shown below:

```
rule(1,"C2-you are stand on vessel",
   ["divergent"],1)
rule(2,"C2-you are stand on vessel",
   ["parallel","speed less"],1)
rule(3,"C2-you are stand on vessel",
   ["parallel","speed equal"],1)
rule(4,"C2-you are stand on vessel",
   ["intersect","speed less"],1)
rule(5,"C2-you are stand on vessel",
   ["intersect","speed equal"],1)
```

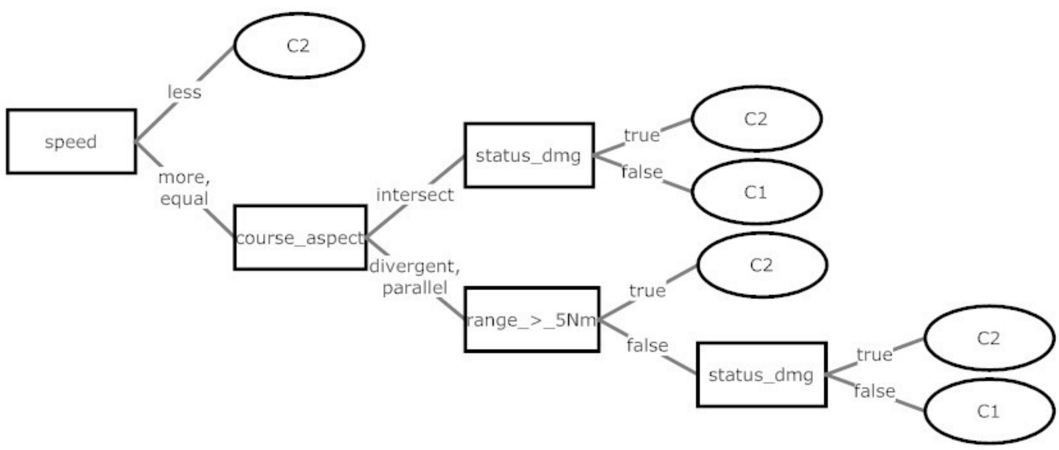

Fig. 1. Decision tree for COLREGS Rule No. 13
Rys. 1. Drzewo decyzyjne dla prawidła 13 MPDM

```
rule(6,"C2-you are stand on vessel",
   ["parallel","speed more",
   "range_more_than_5nm"],1)
rule(7,"C2-you are stand on vessel",
   ["intersect","speed more",
   "range_more_than_5nm"],1)
rule(8,"C2-you are stand on vessel",
   ["parallel","speed more",
   "range_less_than_5nm",
   "status_dmg_true"],1)
rule(9,"C2-you are stand on vessel",
   ["intersect","speed more",
   "range_less_than_5nm",
   "status_dmg_true"],1)
rule(10,"C1-you are give way vessel",
   ["parallel","speed more",
   "range_less_than_5nm",
   "status_dmg_false"],1)
rule(11,"C1-you are give way vessel",
   ["intersect","speed more",
   "range_less_than_5nm",
   "status_dmg_false"],1)
```

By using the verification mechanisms in the system, it was possible to validate the knowledge base, with a positive result. Then the knowledge base was tested in the form of dialogue with the system. In each test the user inputs conditions to the system, which are attribute values. For example, the system inquires about the speed of own ship in relation to target ship, and user points the answer: less, equal, or greater.

The analysis of the set revealed that rule No. 2 and No. 4 can be replaced with a single, simpler rule. The rules were removed from the knowledge base and replaced by a new one:

```
rule(22,"C2-you are stand on vessel",
   ["speed less"],1)
```

Modified knowledge base was re-tested for correctness and test of inference was conducted. Each scenario that was tested gave a positive result.

Machine learning algorithms

In order to run the experiment where machine learning algorithm was used, two sets of data were prepared. Training set and test set were adjusted during the experiment in terms of number of examples, the number and importance of decision attributes and decision classes.

The first example uses the same training set as used in the construction of decision tree above. From the results that are shown in figure 2, it may be noted that, this training set led the algorithm to generate a tree with root node only. All objects in the training set were classified into one class – C2-you are "stand on" vessel. The result obtained in this experiment is obviously unsatisfactory, but it shows the properties of See5.0 algorithm. C5.0 promotes the values of attributes or decision classes that are represented most frequently in the training set.

In order to show the capabilities of the algorithm, the next test with modified training set was conducted. The number of objects in training set was increased up to 85 items by random duplication. The results are shown in figure 3.

It should be noted that the algorithm has made correct classification of objects into classes C1 and C2. Based on synthetic data from set B it has created a decision tree. Tree structure resembles the one obtained by using analytical method described above, but due to the different attribute values distribution, the individual nodes are shifted in relation to each other.

Analysis of results

On the basis of conducted numerical experiments, a comparison of methods of knowledge representation was presented. The results are summarized in table 2.

```
Read 15 cases (4 attributes) from D.data

Decision tree:
 C2-you are stand on vessel (15/3)


Evaluation on training data (15 cases):

            Decision Tree
          ----------------
          Size      Errors

             1    3(20.0%)   <<


           (a)   (b)    <-classified as
          ----  ----
                   3    (a): class C1-you are give way vessel
                  12    (b): class C2-you are stand on vessel
```

Fig. 2. Results of running See5.0 algorithm on training set A
Rys. 2. Wynik działania algorytmu C5.0 – przykład A

```
Read 85 cases (4 attributes) from D.data

Decision tree:

speed in {equal,less}: C2-you are stand on vessel (25)
speed = more:
:...range_>_5NM = true: C2-you are stand on vessel (7)
    range_>_5NM = false:
    :...status_dmg = true: C2-you are stand on vessel (6)
        status_dmg = false:
        :...course_aspect = divergent: C2-you are stand on vessel (3)
            course_aspect in {intersect,
                              parallel}: C1-you are give way vessel (44)


Evaluation on training data (85 cases):

            Decision Tree
          ----------------
          Size      Errors

             5    0( 0.0%)   <<


           (a)   (b)    <-classified as
          ----  ----
            44          (a): class C1-you are give way vessel
                  41    (b): class C2-you are stand on vessel


        Attribute usage:

            100%  speed
             71%  range_>_5NM
             62%  status_dmg
             55%  course_aspect
```

Fig. 3. Results of running See5.0 algorithm on training set B
Rys. 3. Wynik działania algorytmu C5.0 – przykład B

Table 2. Comparison of methods of knowledge representation
Tabela 2. Zbiorcze porównanie analizowanych metod reprezentacji wiedzy

Method	Implementation	Knowledge requires a hierarchy?	Knowledge base updating	Knowledge base verification
Decision tree	for example, in form of nested conditional statements If-then-else; When knowledge is combined with the inference mechanism, such a solution is not compatible with the principle of knowledge bases creation process	yes, depending on the implementation in order to increase the system performance	possible – it requires time-consuming verification or building tree / algorithm from the beginning	difficult, after "mixing" knowledge with the inference algorithm; each time the knowledge is updated an intervention of programmer / knowledge engineer is required
Decision rules (complex)	depending on the number of rules and conditions relatively simple implementation of inference mechanism	yes / no, depends on the quality of the rules	possible	relatively simple verification, implementation of testing mechanisms are required
Decision rules (simple)	simple implementation of the rule base, difficult implementation of inference mechanism	yes, the order of rules and conditions is essential	possible	simple verification of the individual rules, difficult verification of consistency and correctness of the knowledge base as a whole
Induction of trees / rules – Machine learning algorithm	demo version of See5.0 used	no, determined on the basis of calculations	possible – preparation of a new training set and conducting the learning	relatively difficult due to the factor of probability and properties of the algorithm – pruning, classifier reduction

Decision trees allow for relatively easy implementation of knowledge directly into the algorithm of inference. For obvious reasons, complex structures are harder to implement, testing and verifying.

Decision rules method, used in expert systems currently on the widest scale [2] requires implementing complex mechanisms for testing knowledge (knowledge base) and for validation. It plays an important role when rules are being added or modified. Knowledge base with a small number of rules is relatively simple to analyze, while the bases that are constantly updated they might become inconsistent.

For induction of decision trees and rules with machine learning algorithms (See5.0) the result is difficult to predict and would require the in-depth and accurate testing. Thus, the assumption to deliver a properly prepared training set, that covers a larger number of attributes and greater attribute values diversity would provide a complete and effective set of rules or optimal decision tree, is wrong. Decision trees and decision rules generated with machine learning algorithm show that this method of knowledge representation might not be appropriate to store the expert knowledge contained in COLREGS. The uncertainty of the training data, missing attribute values, and pruning of tree branches can lead to the induction of rules that do not overlap with 100% effectiveness the training data. As a result, the inference for the test data might be incorrect.

Conclusions

The described methods of knowledge representation are diverse in their performance and costs associated to their use and testing. Implementation of the COLREGS rules as the knowledge base is a complex process. Each of the methods of knowledge representation requires the involvement of a knowledge engineer. Discussed solutions allow for knowledge modifications, but require the implementation of verification mechanisms. There is also a relation between different methods that a relatively simple implementation of knowledge could force to implement complex inference and verification mechanisms and vice versa.

Due to the significant dynamics, variability and ambiguity of navigational situations, which are observed during the meeting of two or more vessels (e.g. the impact of weather and sea conditions on vessels, undetermined values of attributes, etc.), it should be considered to analyze to build MPDM knowledge base with using the fuzzy sets [13]. It is planned to analyze the representation of the described methods for the use of continuous attributes (bearings, speed of vessels, etc.). It is also essential to consider and to conduct testing of other methods of knowledge representation, including rough sets.

References

1. http://encyklopedia.pwn.pl/haslo/3995573/wiedza.html (access 15.02.2012).
2. MULAWKA J.J.: Systemy ekspertowe. Wydawnictwa Naukowo-Techniczne, Warszawa 1996.
3. NĘCKA E., ORZECHOWSKI J., SZYMURA B.: Psychologia poznawcza. PWN, Warszawa 2006.
4. PIETRZYKOWSKI Z., URIASZ J.: Knowledge representation in a ship's navigational decision support system. In Adam Weintrit (ed.), Marine Navigation and Safety of Sea Transportation, Gdynia 2009, 45–52.
5. RUTKOWSKI L.: Metody i techniki sztucznej inteligencji. Wydawnictwo Naukowe PWN, Warszawa 2009.
6. PIETRZYKOWSKI Z., MAGAJ J., CHOMSKI J.: A navigational decision support system for sea-going ships. Measurement Automation and Monitoring, 10, 2009, 860–863.
7. WOŁEJSZA P.: An Algorithm of an Anti-collision Manoeuvre. Międzynarodowa Konferencja Naukowo-Techniczna Inżynieria Ruchu Morskiego, Świnoujście 2005.
8. RYMARZ W.: Podręcznik Międzynarodowego Prawa Drogi Morskiej. Gdynia 1995.
9. KWIATKOWSKA A.M.: Systemy wspomagania decyzji: jak korzystać z wiedzy i informacji w praktyce. PWN, Warszawa 2007.
10. QUINLAN J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.
11. KOTSIANTIS S.B.: Supervised Machine Learning: A Review of Classification Techniques. Informatica 31(2007).
12. NIEDERLIŃSKI A.: Regułowo-modelowe systemy ekspertowe rmse. Wydawnictwo Pracowni Komputerowej Jacka Skalmierskiego, Gliwice 2006.
13. PIEGAT A.: Modelowanie i sterowanie rozmyte. Akademicka Oficyna Wydawnicza Exit, Warszawa 2003.