Advanced Decision-Making Systems in Future Avionics: Automatic Target Recognition Example

Edmund G. Zelnio
Air Force Research Lab
Sensors Directorate
WP-AFB, OH 45433
937 255-4949 x4211
ezelnio@mbvlab.wpafb.af.mil

Abstract—In future avionics systems, there will be a stronger and stronger pull for more automation of decision-making functions. The nature of these decision-making functions will be of such complexity that exhaustive testing of all possible decisions will be impossible. Further, the complexity will also preclude exhaustive "training" of the decision-making systems such as the training required of learning-based paradigms like neural networks. The primary thesis of this paper will advocate a change of focus in training and testing procedures for advanced decision-making systems such as automatic target recognizers (ATR's). Whereas currently, the predominant state of the art in ATR testing is preoccupied on making sure that the training and testing data come from the same population. This common procedure helps promote excellent results but does not provide insight into performance of the system under real world conditions. Rather, the focus should be on whether the testing is representative of the real world or whether it even scales to the complexity of the real world situations. Testing performed in this way will result in different approaches to decision-making systems such as ATR and, it is believed, will accelerate progress in ATR and other decision-making functions.

TABLE OF CONTENTS

1. INTRODUCTION
2. COMPLEXITY OF PROBLEM SPACE
3. HOW DOES IT WORK?
4. HOW DOES THE SYSTEM PERFORM ONCE KNOWLEDGE IS OBTAINED (ON-LINE)?
5. HOW QUICKLY, EFFICIENTLY CAN KNOWLEDGE BE ADDED TO THE SYSTEM (OFF-LINE COMPONENT)?
6. EVALUATING COMPLEX DECISION-MAKING SYSTEMS
7. CONCLUSIONS

1. INTRODUCTION

The avionics systems of the future are automating more and more decision-making functions as the workload of the pilot becomes greater, and the trend towards unmanned air vehicles continues. Many of the decision-making functions are of extreme complexity, and therefore, preclude exhaustive testing of all possible conditions for which the systems are designed to operate.

There is reason for concern since without careful design and test, systems will not operate correctly. Since these decision-making systems often are in the critical path of life or death decisions, significant attention needs to be focused on the issue of scientific evaluation of these complex systems.

This paper examines the issues associated with evaluation of complex decision-making systems by focusing on the issues that surface in automatic target recognition systems. These issues are key issues that must be faced when developing and testing complex decision-making systems. In particular, this paper addresses the evaluation of complex decision-making systems by asking the following questions:

How does it work?

Although one would like to be able to just do evaluation from a black box perspective (i.e., just worry about I/O), knowing how the decision-making system works will provide the evaluator with necessary constraints to help effectively evaluate the system. As will be seen, the complexity of the problem space is much too large to sample exhaustively. Hence, smart sampling will be helped by knowledge of how the system performs its decision-making functions.

How does the system perform once knowledge is obtained (on-line component)?

A key aspect of evaluating on-line system performance is to understand and even control the knowledge obtained via the off-line or training portion of the system. Again, this need derives from the complexity of the problem space and the resulting realization that the space will be exceedingly sparsely sampled. Knowing where the space is sampled in both training and testing is essential to drawing meaningful conclusions in the evaluation process.

How quickly, efficiently can knowledge be added to the system (off-line component)?
For complex decision-making systems, there is always a need to add more knowledge as the situation or mission changes. The continued value of these systems will be closely tied to the ease by which knowledge can be added, and hence, the knowledge acquisition process becomes an important evaluation factor.

Many of the challenging issues associated with evaluation of these systems starts with the complexity of the problem space.

2. Complexity of Problem Space

The complexity of the problem space is due to the extreme complexity of the real world in which the decision system must operate. For example, if one considers just the complexity of the target signature space for a typical ATR problem, the development and evaluation challenges due to this complexity can be illustrated.

The complexity of the problem space can be characterized by considering the number of possible target signatures that the ATR system may have to recognize. For example, for a 20 class problem using a synthetic aperture radar (SAR) sensor, the numbers of signatures are enormous. If one looks at modest estimates of variation due to target considerations — type, configuration, articulation, obscuration, layover, and netting; and due to sensor acquisition geometry considerations — aspect angle and depression angle, the number of signatures are shown in Figure 1 below. For other sensor types, one could easily get significantly larger numbers as SAR has constant scale and is active so its amplitude does not vary with environmental conditions as do passive sensor types.

<table>
<thead>
<tr>
<th>Components</th>
<th>Variations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Type</td>
<td>20</td>
</tr>
<tr>
<td>Target Aspect</td>
<td>72</td>
</tr>
<tr>
<td>Depression Angle</td>
<td>5</td>
</tr>
<tr>
<td>Articulation (DoF)</td>
<td>36</td>
</tr>
<tr>
<td>Configuration (4 binary)</td>
<td>16</td>
</tr>
<tr>
<td>Obscuration</td>
<td>400</td>
</tr>
<tr>
<td>Layover</td>
<td>20</td>
</tr>
<tr>
<td>Netting</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>165,888,000,000=1.6*10^{11}</strong></td>
</tr>
</tbody>
</table>

Figure 1. Combinatorics Associated with SAR ATR

The problem complexity described above has just considered the complexity of the target. Of course, the complexity of the background also comes into play and must also be considered. It is clear, however, that only considering the combinatorics of the targets are sufficient to illustrate the extreme complexity of the problem space facing automated decision-making systems.

3. How Does It Work?

Understanding how the algorithm works is important to knowing how to evaluate the system. An important question, however, is what knowledge is important to the understanding of the particular decision-making system. This question is answered by first categorizing algorithms into two types – model-based and learning-based algorithms.

There are several ways to categorize algorithms. One method that is quite useful is to characterize how the algorithm obtains its knowledge about the problem space. Fundamentally, there are two methods[1]. One method called exemplar-based or learning-based obtains its knowledge implicitly via examples. These algorithms do not assume any knowledge of the domain except what can be ascertained from the ensemble of examples. The other method called model-based obtains its knowledge explicitly via models.

**Learning-based Algorithms**

Based on the fact that learning-based algorithms obtain their knowledge from examples, they are capable of interpolating between sample points or examples but are not capable of extrapolation. Hence, they are only capable of recognizing things that they have seen before or close to what they have seen before (i.e., they can generalize).

Learning-based algorithms such as many types of neural nets have become very popular since the decision process is learned from examples rather than from any understanding of the underlying sensing physics. Under conditions where training data is representative of the conditions that an algorithm will see in the real world situation, algorithms can have very good performance. Performance necessarily degrades, however, as testing conditions deviate from training conditions.

**Model-based Algorithms**

Model-based algorithms theoretically can extrapolate beyond situations where they have been given data assuming that the models accurately model the conditions where data is missing. The advantage of model-based systems[2] is that the underlying physical processes are modeled which allows the algorithm to extend to conditions for which the algorithm has not been trained, but for which the system is modeled. First principle-based models should be capable of extrapolation beyond data. However, these models must be of high fidelity to support fine discrimination; hence, validation is a key issue with these type systems. The Kalman filter is an excellent example of a model-based system where the dynamical models of the system are embedded into the filter.

Most practical systems will have elements of both model-based and learning-based systems and, therefore, will be hybrid. In order to perform evaluation in the high dimensional problem space, it will be useful to understand which parts of the system are model-based and which parts
are learning based. This thought will be developed later in this paper.

4. **How Does the System Perform Once Knowledge Is Obtained (On-Line Component)?**

In order to test the system, one is faced with a number of considerations that complicate the evaluation of complex decision-making systems such as ATR. Complexity of the problem space and how algorithms work have been discussed. Important considerations, which also drive any evaluation of decision-making systems such as ATR, include the statistical nature of decision systems, training and testing over part of the problem space, and the coupling of robustness vs. selectivity or, for ATR, probability of detection vs. probability of false alarm.

**Statistical Nature of Decision Systems**

The statistical nature of decision systems requires that a number of trials be performed before anything definitive can be said about system performance. Hence, a single measurement is really meaningless for these automated decision-making systems. Evaluation of statistical processes requires a stated hypothesis with a number of trials sufficient to prove or disprove the hypothesis with a desired confidence. Problem spaces of the complexity described above present significant challenges from a statistical sampling perspective. The experimental design of the evaluation must be as economical as possible in sampling the space yet must also be able to make statistical inferences with reasonable confidence.

**Development and Testing over Part of Space**

Evaluation of the system will necessarily be performed on a very small subset of the entire space of possibilities. It is important to carefully consider the subset of training and testing in the design of the evaluation. Since the space is very large and the data will be necessarily limited, one must be sure that some of the training and test sets are from disjoint sets of the problem space. There is a dangerous tendency to split training and test sets such that they come from the same statistical distribution sampled over the same conditions. This split ensures that about any algorithm will work under the conditions specified in the test; however, this test tells us nothing about how the technique will work in the real world. In order to say something about real world performance, one needs to understand how the system performs outside the conditions for which training data exists. Since practically, one can never have enough data to cover all the conditions of the problem space. It turns out that testing under conditions that are not in the training set is the exception rather than the rule — certainly in the ATR community. This practice needs to be reformed.

**Robustness vs. Selectivity or Pd vs. Pf**

There is an intrinsic tie between robustness of a system and its selectivity, which is also related to the probability of detection of the system and its false alarm rate. There is danger in focusing on one aspect of the problem without paying attention to the other. For example, the ATR community has, at times, focused on false alarm rate to the exclusion of target variability. This over focus on false alarm rate has led to ATR's which were trained on a single serial number of a target (e.g., one instance of a T72) rather than the total target class. This has led to overly optimistic false alarm rates since the target filter, in effect, was overly restrictive — only recognizing a single T72 tank rather than the class of T72 tanks.

This relationship between robustness and selectivity must be considered when developing the evaluation experiment. Again for ATR, one must sample both target and background conditions. Undersampling the target can result in a “target filter” that is too narrow, which will result in an ATR that misses the target because its knowledge of the target is incomplete. Undersampling the background will give optimistic estimates of false alarm rate since the backgrounds in the real world will be much more varied than tested and, therefore, could result in more false alarms. In any event, the evaluation of automated decision-making systems must be careful to sample the space with both robustness and selectivity in mind.

5. **How Quickly, Efficiently Can Knowledge Be Added to the System (Off-Line Component)?**

![Figure 2. Options for Knowledge Acquisition to Handle Large Parameter Variation of Problem Space](image-url)
Another consideration that is important for automated decision-making systems such as ATR is the ability to add new knowledge to the system. Often the decisions concern situations or conditions that change. New knowledge needs to be added. The ease, timeliness, and cost of adding knowledge is a key issue that needs to be evaluated in order to determine the maturity of the decision-making system for fielding.

Data and Models

Given that the complexity of the space is a central issue in developing decision-making systems based on sensor data, one needs to decide on how to obtain the knowledge of the space for the decision-making system development. Fundamentally, two approaches are available - data and models. In order to obtain the knowledge over the large number of parameters representing the problem space, one has three options - collect data, model, or do neither represented as punt in the illustration in Figure 2 above.

Perhaps ideally, one would gather data over all conditions; however, for many problems this would be prohibitively expensive and even intractable. Since we are trying to minimize time and cost in adding knowledge to the system, exhaustive data collection does not seem to be the answer. Modeling the target variations represents an attractive alternative to data collection; however, the issue of model fidelity is key as it limits the level of discrimination that one can attain for an ATR using models as the basis for its target knowledge. Many times, however, one is faced with the reality that some parametric variations are not easily handled by systematic data collection or by models. For example, real world variations due to mud on tanks or sensor variations due to spurious sensor errors are likely variations that the ATR will face but will probably not be represented by the ATR.

The key issue here is that one should explicitly determine how the problem space parameter variation would be handled from a knowledge acquisition standpoint. This requires a combined strategy that takes into account explicit strategies for handling each set of parameters. For example, an ATR approach must be insensitive to the punt parameters or potentially suffer unacceptable performance degradation. Knowledge acquisition strategies clearly must be consistent with the algorithm decision-making process.

Many perception and sensing systems are trying to sense something that is, at best, non-cooperative and, often, is adversarial. Hence, new targets and signature modification techniques are realities that a system will need to cope with. This critical problem centers, again, around the dimensionality of the space. It is quite difficult to collect enough data to sample the space to use data as the fundamental basis for adding new targets. Hence, modeling can play a large role in updating the recognition system as new targets emerge. Modeling is essential when dealing with adding denied target signatures to a recognition system since remote sensing is the primary method of obtaining signature information.


Many issues have been discussed that deal with evaluation of complex decision-making systems. These center around the fact that these systems must make inferences and decisions in real world conditions which have an enormous problem space associated with them. One of the advantages that evaluators can and should utilize in evaluating automated complex systems is the knowledge of how the system makes decisions. In order to explore how this can be an advantage, let’s look at Figure 3 below.

![Figure 3. Partitioning of the Problem Space](attachment://figure3.png)

The problem space is partitioned based on knowledge of the decision-making system being evaluated. In particular, the space is partitioned into three regions: regions for which there is data, regions that are modeled, and the remaining regions for which neither data nor a model representation exists[3]. Recall that we are required to evaluate performance or to determine whether the system will work in all regions of the problem space. Given this partition, the evaluation can first determine whether the system works in the regions where it has been trained or development data. This is called the validation region since we are validating that the system works in regions where it has seen data before. Next, given that the system in some way models a subset of the problem space, the evaluation next samples this subspace to determine performance in this region. This is called the extendibility region since it extends the system performance beyond the data but within the parametric span of the underlying models of the system. Finally, the last partition is the region where neither data nor models represent knowledge in the ATR algorithm. This region is termed the robustness region since the algorithm’s performance must be robust to deviations from its internal target representations whether they are derived from data or models.

Systematically exploring each region can give insight into the operation of the system and help characterize its performance based on knowing how the system functions.
This methodology can be taken further by exploring parameterizations of the space based on the ATR algorithm. For example, if one knows that the algorithm handles articulations of targets with a particular methodology, one can test whether that methodology is effective by testing a few sample points along the articulation dimension rather than having to sample this dimension exhaustively. The sampling strategy would also be influenced by whether this was an exemplar-based or model-based part of the system. Further, if it were exemplar-based, the evaluation samples would be chosen based on the particular samples used in the training or exemplar set and whether the test concerned the validation or robustness region of the problem space.

It is important to note that the evaluation is structured to test the full dimensionality of the problem space and is not limited to the area of the space where training data was used. This methodology, as mentioned earlier, is critically important to the evaluation of complex decision-making systems. Evaluations should be specified in terms of the parameterization of the problem space (e.g., target types, conditions, background types, etc.) rather than what is contained in a training set.

Measured data is not the only source for evaluating these systems. Judicious use of synthetic data and theoretical performance predictions provide the evaluator with additional tools to explore the large dimensionality of the problem space. Judicious, here, has two implications. The first one is that both synthetic data/simulations and theory should be validated before one puts too much trust in the resulting evaluation. The second, however, gets back to understanding the operation of the system. If the evaluator understands how the system operates, then the judicious use of synthetic data and theory can be based on knowledge of whether the particular model or data limitation pertains to the particular sensitivity of the algorithm. This methodology acknowledges that models or theories are only approximations. Hence, the test can be designed such that the algorithm will be insensitive to the modeling errors under the conditions of the test and, therefore, the models may be quite useful in the evaluation of a particular dimension of the problem space.

7. Conclusions

The evaluation of complex decision-making systems is a complex process itself. It is critically important as more and more systems become automated and rely on algorithms to make decisions rather than humans. The reasons for this complexity are illustrated by examples drawn from the automatic target recognition area. The central reason for the extreme complexity is the high dimensionality of the problem space. It is too large to sample effectively during testing. Hence, smart sampling is a necessity for the evaluator. In order to perform smart sampling, several issues are discussed including: the statistical nature of decision systems, the need to carefully partition the training and testing evaluation sets, the relationship of robustness and selectivity of algorithms, and, very importantly, the knowledge of how the decision-making system works. It is argued that knowledge of how the decision system works in terms of whether its reasoning strategies were either exemplar-based or model-based could be quite helpful in determining smart sampling strategies for evaluating with measured data or, even, where simulation and theoretical predictions may be useful. Finally, it is pointed out that an important part of any decision-making system is the off-line component of the system or that component that acquires knowledge for the on-line component. Since the scenarios, conditions, and situations often change for these systems, the ability to efficiently and effectively add knowledge to the system is a key element that also requires careful evaluation in order to assess the true usefulness of these systems.

REFERENCES


Edmund G. Zelnio graduated from Bradley University, Peoria, IL, in 1975, and has pursued doctoral studies at The Ohio State University, Columbus, OH. He has had a 22-year career with the Air Force Research Laboratory, Wright Patterson AFB, OH, where he has spent 20 years working on automatic target recognition development. He is currently the technical director of the Sensors ATR Division of the Sensors Directorate, Air Force Research Laboratory. He also serves in an advisory capacity to the Department of Defense and the Intelligence community in the areas of ATR and information technology. He has served as the Chair of the ATR subarea for the Defense Technology Area Plan, which integrates the Air Force, Navy, Army, DARPA, and BMDO technology plans. He has guided development of prototype ATR systems for both air-to-air and air-to-ground applications using both electro-optical and radar sensors. His primary focus has been the development of model-based ATR technology for synthetic aperture radar systems. Mr. Zelnio is a member of the IEEE and the Automatic Target Recognition Working Group, a joint industry-university-government group that serves the collective interests of the ATR community.