45 International Scientific Colloquium
Ilmenau Technical University
October 04 - 06, 2000

A.J. Gonzalez / M. Murillo / R. Knauf

Vaslidating Human Behavior Models

Abstract

Human behavior models (hbm's) have been used by the military training community to assist in tactical training tasks. These models represent typical human behavior in tactical situations such as seen in battle. They can serve to automate the presence of enemy as well as friendly units, and thus save the manpower effort required to populate and control these entities. More recently, they have been used as part of larger aggregate units in constructive simulations to provide greater accuracy and realism to these simulations. Validation of these models has not yet been a research issue because of the community's current preoccupation with building them. However, this is currently gaining importance, as developers transition from a research to a deployment mode. The dynamic nature of these models and the need to use experts to judge their validity introduce great difficulty in validating hbm's. This paper discusses some conceptual approaches to this problem. These are based on comparing the behavior of the model to that of an expert, while the latter behaves normally in a simulated environment, under the same conditions as perceived by the model.

1 Introduction

Human behavioral models, known in the military simulation and training community as *Computer Generated Forces* (CGF), have been successfully incorporated in several significant simulation training systems. However, there is a serious need on the part of these users to be able to validate the behaviors demonstrated in order to ensure a sound training process.

Because of its intuition and popularity, most of the early systems that implemented CGF's were based on rules. Golovcsenko [Gol87] discusses some U.S. Air Force prototype systems that exhibit certain amount of autonomy in the simulated agents.

Another popular technique used in CGF systems has been *Finite State Machines* (FSMs). FSM's have been used to implement goals or desired states in the behavior of the CGF ([Dean96]). These states are represented as C-Language functions. These systems employ FSMs as the representational paradigm. The knowledge found on FSM's does not necessarily aggregate all the related tasks, actions, and things to look out for in a self-contained module. This makes their formalization somewhat difficult from the conceptual standpoint. Some FSM-based systems allow for the control of one entity by more than one FSM at the same time. This can be dangerous in that an incorrect behavior can be easily displayed.

Other alternative representation and reasoning paradigms exist. Techniques such as model-based, constraint-based or case-based reasoning, although promising in some respects, (see [Bor77], [Cas91], [CM92]) not easily capture the heuristics involved in tactical behavior representation. Thus, they are not considered "natural" for this type of applications.

Another common approach to the CGF problem has been to use blackboard architectures to represent and organize the knowledge ([CS86]).

Yet another approach has come from cognitive science researchers ([WCP92], [ZZR89], [Za89]).These efforts do not directly address CGF's, but rather, the closely related problem of cognitive modeling of the human decision-making process.

As can be seen, there are several means of representing the knowledge required to model human behavior as it applies to tactics. The problem remains how to validate the behavior models in a way that makes all the different representational paradigms transparent. The next section discusses some conceptual approaches to the problem.

2 State of the Art in Validation Techniques for Human Behavior Models

Birta and Ízmizrak ([BI96]) propose an approach for behavioral validation of simulation models, which is intended to perform automatic validation. The key element of their software environment is a *validation knowledge base* (VKB), whose objective is to comprehensively identify the expected behavior of the simulation model. Based on the information in the VKB, it defines a set of experiments to be performed by the simulation model. Finally, the results of the execution are compared against the expected behavior in the VKB to determine the validity of the model. Their framework for developing the VKB is formulated around the concept of dynamic behavior and it is represented by means of an abstraction called *dynamic object*. A dynamic object O is an ordered pair of vectors X and Y, i.e., $O = (X, Y)$, where X corresponds to the input of the dynamic object and Y corresponds to the output of the object. The fundamental property of the dynamic object is "its ability to generate (exhibit) behavior over some prescribed time interval" ([BI96], p. 81). It's assumed a causal relationship exists between the vectors X and Y. The VKB should contain all possible instances of a dynamic object. These instances or elements are described in terms of their attributes, i.e., their input and output vectors. This characterization is presented as the union of three disjoint sets of relationships.

Caughlin ([Cau95]) denotes the lack of efficient techniques to compare a simulation with the real situation that is being modeled, or to compare two different models of the same situation. His proposal is directed to both verification and validation of models and simulations. However, only the validation process is addressed in this paper. His idea is to abstract the original model and to reduce its complexity by aggregating the details into a more general module. Instead of analyzing the different parts of the model and then trying to integrate the results, Caughlin proposes to analyze the model as a whole and determine whether it meets the expected behavior according to the real phenomenon under study. This approach consists in reducing the order (to abstract) of the model representation by means of what he calls *reduced order metamodels*. These metamodels would be used as black boxes that approximate the causal time dependent behavior represented by the simulation. As he states, "as an abstraction, a metamodel is a projection of the model onto a subspace defined by new constraints or regions of interest" ([Cau95], p. 1408). According to Caughlin, this approach is cost effective, timely and objective ([Cau95], p. 1412).

Page and Canova ([PC97]) discuss some mechanisms and research directions that need to be applied to confront the differences that the verification and validation processes for advanced distributed simulation (ADS) present. One of the most important ADS is the Aggregate Level Simulation Protocol (ALSP) Joint Training Confederation (JTC). JTC is a collection of training simulations that supports joint training at the command and battle staff levels during major. The authors describe in their paper the mechanisms used for the verification, validation and accreditation processes of ADS. They present the development process of the JTC and they emphasize the importance of validation in each stage of this process as well as a general validation of the whole system.

The approach proposed by Hone and Moulding ([HM98]) consists of improving the understanding of the application domain in which the model (called by the authors *Synthetic Environment*) is to be employed. This is done instead of focussing only in the problem domain being modeled as most of the validation and development techniques for simulations currently do. Their objectives are to complement existing development approaches for synthetic environments (SE) with application domain modeling, and to define a suitable set of applications-oriented models which are relevant to the military training and to the military equipment procurement. The authors argue that these models could be used as the basis for the verification and validation processes. According to the authors, this new approach would provide stronger validation capabilities to the development process of synthetic environments. At the same time, they suggest carrying out the validation process as early as possible in the development process in order to minimize possible costs and risks.

3 Validation Through Observation of Agent Behavior

Since these models are by definition designed to simulate human behavior, it becomes clear that they must be compared to actual human behavior. Validation of the more traditional expert systems is really no different, as these attempt to model the problem solving ability of human experts. Expert systems have traditionally been validated using a suite of test cases whose solution by human domain experts is known ahead of time. However, these tests are generally static in nature – they provide the system with a set of inputs and obtain its response, then check it against the expert's response to the same inputs. Time is typically not part of the equation, unless it is already implicitly incorporated into the inputs (i.e., one input could represent a compilation of the history of one input variable). The techniques suggested by Abel ([AG97]) provide effective and efficient means of generating good test cases based on the validation criteria specified for the intelligent system. The Turing Test approach proposed by Knauf ([KJGP98]) is a promising way to incorporate the expert's opinion in a methodical fashion for time-independent problems and their time-independent solutions.

Validating human behavioral models, on the other hand, requires that time be explicitly included in the expression of the tactical behavior. Such behavior not only has to be correct, but also timely. Reacting to inputs correctly, but belatedly can be the difference between life and death for the decision-maker in a battlefield. Furthermore, tactical behavior is usually composed of a sequence of decisions that are made as the situation develops interactively. Such developments are generally unpredictable and, therefore, it becomes nearly impossible to provide test inputs for them dynamically.

One approach could be to "discretize" the scenarios being presented to the expert validators by decomposing them into highly limited situations that would take the time out of the equation. The experts would then have to visualize the scenario and decide whether the model's decision is the appropriate one. Several of these scenarios would have to be strung together in sequence in order to make the entire test scenario meaningful. This would be an artificial setting, and the expert may have difficulty in visualizing the actual situation when presented in this fashion. It is possible that he would do something different if he were actually facing that situation in a more realistic setting. Furthermore, interaction would not be possible. Therefore, this is generally not considered acceptable as a mainstream solution.

Alternatively, from a conceptual standpoint, validation can be executed for these types of models by comparing the performance of the expert with that of an *autonomous intelligent platform* (or AIP) controlled by the model being validated. The AIP and the human would both be subjected to the same initial inputs, mission, environment, etc. The performance of each (the AIP and the validating expert) can be represented as a sequence of data points for the observable variables (e.g., x, y, t and others) over a period of time. Overlaying one on top of the other may provide some indication of validity for the model's performance if they generally agree. A complete match between the expert's performance and the AIP's behavior would certainly justify validation of the model that controls the AIP. Realistically, however, a significant amount of deviation may exist between the two performance records. While some deviations may be indicative of a serious discrepancy in behaviors, others may simply be a different and equally appropriate way of achieving the same goal. Thus, expert input may be necessary to determine what is correct and what is not correct. An intelligent system could be developed to perform this last task of determining what is an acceptable deviation and what is not, but it would also ultimately have to be validated itself, and that would ultimately require human expertise.

Furthermore, due to the interactive nature of the model and the domain, the model being validated may make a different decision from what was made by the validating expert which, although acceptable, progresses into a different scenario. Consequently, the two performance records could no longer be adequately compared, as their situations may have diverged significantly enough to make them not relevant to each other. This is referred to as loss of synchronism between the model and the human validators.

A third approach would be to have the validation expert(s) observing in real time the performance of the model in a specific scenario. He (they) would then be able to designate the observed decisions by the model as acceptable or unacceptable while watching it in real time. This approach maintains synchronism between the model and the experts. However, it is considered a very intense exercise for the validation experts, as they would have to continuously attend to the scenario and compare their decision to the model's decision. This would severely limit the number of exercises that would be possible to execute for validation.

In reality, none of the above techniques provides us with an effective and efficient means to validate the performance of human behavioral models. They are cumbersome, unnatural and/or difficult to implement.

4 Validation by Comparison with Expert Performance in a Simulation

One promising alternative is to have the expert validator perform the task, rather than merely monitor its performance by the model. As he executes it, he does so in parallel with an AIP controlled by the model in real time, and within the same virtual environment. The two are faced with the same conditions (at least initially), and mission. A comparison system called the *Difference Analysis Engine*, or DAE, continuously checks for discrepancies between the two. The absence of discrepancies indicates correct behavior by the model-controlled AIP. However, not all discrepancies are errors, as in many instances there may be more than one acceptable way of doing something. As in the prior discussion, this is something to be determined off-line by either an expert validator panel, or some sort of automated system. In either case, this is beyond the scope of the work described here.

The value of the DAE is the maintenance of synchronism between the AIP and the human. This synchronism allows the monitoring of discrepancies to continue to be viable throughout the entire. Loss of synchronism means that the two behaviors will diverge from each other. This could be the result of an incorrect decision by the model, or because of a correct, but different decision by the expert. Once divergence sets in, it is not meaningful to continue the discrepancy monitoring process, as the two entities will not be in the same situation. It is also not realistic to expect the entities' behaviors to re-converge if left to their own devices. The DAE not only reports and logs all discrepancies, whether important or otherwise, but it also ensures that the AIP will be re-synchronized with the human after a discrepancy has occurred that causes a divergence.

The re-synchronization process depends on the modeling paradigm being utilized by the model. Here we will base our discussion upon the Context-based Reasoning (CxBR) paradigm developed by Gonzalez and Ahlers ([GA95]). Modeling with CxBR assumes that the AIP being modeled is always under the control of a context that contains the required knowledge to properly control the AIP through its current situation. The model is able to recognize the correct situation with the help of the current context, which helps by pruning the search space for other contexts to which to transition. Therefore, life for a CxBR-based model is a sequence of transitions between contexts that describe the developing situations the AIP faces. Each context is correct for a specific situation the AIP faces, and the context knows when to expire and activate (transition to) a new context when the situation changes so as to render it invalid.

The DAE's task is to recognize the context in which the human operates at all times during the simulation. It then assures that the context of the model controlling the AIP be the same one. As long as they are in the same context, synchronism is likely. It is possible that the human and the AIP will not always be completely synchronized while still being in the same context. This is because the context does not necessarily control the AIP's location and speed to the fine granular level of detail necessary to predict the human's exact location. Therefore, some continuous relocation of the model to overlay directly above the human may be necessary. However, this is not seen as a problem, as such small discrepancies are typically not significant.

The key to this approach is the correct interpretation of the human's context. A technique known as *Template-based Interpretation* (TBI) can be used towards this goal. TBI creates templates that are descriptive of the observable low-level actions performed by a human under that context. Low-level actions are defined to be acts that are simple in nature, and of rather short duration. A sequence or collection of these low-level actions can indicate to an observer the intentions of the human decision-maker. A template is composed of attributes – yes/no checks that indicate whether that action has been observed to have taken place or not. TBI then becomes a competition among several relevant templates for the right to identify the human's actions. The template most completely filled at any one time is the one most likely to represent the human's behavior. Thus, TBI is a continuously-running process.

When a discrepancy is logged by the DAE, it then proceeds to make a synchronization decision about the AIP's context, in light of the human's. The following are the decisions that can be made:

- If the contexts are still the same, then it merely teleports the AIP's location to that of the human and continues.
- If the contexts are not the same, it then forcedly shifts the context of the model to agree with that of the human, teleports the AIP to the location, heading and speed of the human, and continues.
- If it does not recognize the human's context, it teleports the AIP to the human's location, speed and heading, and proceeds to continue to do this repeatedly until the context is recognized and the model's context can be shifted.

Lastly, the validation process becomes a compilation of all discrepancies logged and their individual importance. This is an important area of research not addressed by the DAE. However, the DAE provides an infrastructure that permits the collection and logging of all discrepancies in a relatively simple manner. It can also serve to format and display the discrepancies for ease of analysis by the validators.

5 Summary and Conclusion

It is clear that validation of human behavioral models introduce a new level of difficulty in the validation of intelligent systems. Conventional validation techniques, such as the ones used for expert systems, may not be effective for such a task. A promising alternative exists with the concept of validating a model through real-time comparison with expert behavior. While this concept is relatively immature and requires significant further investigation, it represents a very viable approach to this very difficult problem of validating human behavior models.

References

[Gol87] I.V. Golovcsenko, "Applications of Artificial Intelligence in Military Training Simulations", Master's Degree Research Report, University of Central Florida, 1987.

[Dean96] C.J. Dean, "Semantic Correlation of Behavior for the Interoperability of Heterogeneous Simulations", Master's Thesis, Department of Electrical and Computer Engineering, University of Central Florida, May 1996.

[Bor77] A. Borning, "ThingLab - An Object-oriented System for Building Simulations Using Constraints", *ACM Transactions on Programming, Languages and Systems*, 3(4), pp. 353-387, 1977.

[Cas91] D. Castillo, "Toward a Paradigm for Modelling and Simulating Intelligent Agents", Doctoral Dissertation, University of Central Florida, December 1991.

[CM92] N. Catsimpoolas, J. Marti, "Scripting Highly Autonomous Behavior Using Case-based Reasoning", *Proc. of the 25th Annual Simulation Symposium*, Orlando, FL, April 1992, pp. 13-19, 1992.

[CS86] Y.Y. Chu, D. Shane, "Intelligent Opponent Simulation for Tactical Decision Making", Report PFTR-1120-11/86, sponsored by the Office of Naval Research and the Naval Training Systems Center, 1986.

[WCP92] M.Z. Weiland, B. Cooke, B. Peterson, "Designing and Implementing Decision Aids for a Complex Environment Using Goal Hierarchies", *Proc. of the Human Factors Society 36th Annual Meeting*, 1992.

[ZZR89] M.C. Zubritsky, W.W. Zachary, J.M. Ryder, "Constructing and Applying Cognitive Models to Mission Management Problems in Air Anti-Submarine Warfare", *Proc. of the Human Factors Society 33rd Annual Meeting*, pp. 129-133, 1989.

[Za89] W.W. Zachary, "A Context-based Model of Attention Switching in Computer Human Interaction Domain", *Proc. of the Human Factors Society 33rd Annual Meeting*, pp. 286-290, 1989.

[BI96] L.G. Birta, F.N. Ízmizrak, "A Knowledge-Based Approach for the Validation of Simulation Models: The Foundation", *ACM Transactions on Modeling and Computer Simulation*, 6(1), January 1996.

[Cau95] D. Caughlin, "Verification, Validation, and Accreditation (VV&A) of Models and Simulations through Reduced Order Metamodels", *Proc. of the 1995 Winter Simulation Conference*, 1995.

[PC97] E.H. Page, B.S. Canova, "A Case Study of Verification, Validation, and Accreditation for Advanced Distributed Simulation", *ACM Transactions on Modeling and Computer Simulation*, 7(3), July 1997.

[HM98] G. Hone, M. Moulding, "Application Domain Modeling to Support the Verification and Validation of Synthetic Environments", *Spring Simulation Interoperability Workshop*, Orlando, FL, March, 1998.

[AG97] T. Abel, A.J. Gonzalez, "Enlarging a Second Bottleneck: A Criteria-based Approach to Manage Expert System Validation Based on Test Cases", *Proc. of the 1997 Florida Artificial Intelligence Research Symposium*, Daytona Beach, FL, May 1997.

[KJGP98] R. Knauf, K.P. Jantke, A.J. Gonzalez, I. Philippow, "Fundamental Considerations for Competence Assessment for Validation", b*Proc. of the 11th International Florida Artificial Intelligence Research Society Conference*, Sanibel Island, FL, May 1998, pp. 457-461, 1998.

[GA95] A.J. Gonzalez, R.H. Ahlers, "Context-based Representation of Intelligent Behavior in Simulated Opponents", *Proc. of the 5th Conference in Computer Generated Forces and Behavior Representation*, Orlando, FL, May 1995.

Author Information:

Prof. Avelino J. Gonzalez Ph.D., P.E. and Maureen Murillo
University of Central Florida, School of Electrical Engineering and Computer Science
PO Box 16 24 50, Orlando, FL 32816-2450, USA
phone: ++01-407-823-5027, fax: ++01-407-823-5835, E-mail: *gonzalez@pegasus.cc.ucf.edu*

Dr.-Ing. Rainer Knauf
Ilmenau Technical University, Faculty of Computer Science and Automation
PO Box 10 05 65, 98684 Ilmenau, Germany
phone: ++49-3677-69-1445, fax: ++49-3677-69-1665, E-mail: $rainer.knauf@theoinf.tu - ilmenau.de$