CHAPTER XX

Interfacing and Validating Models of the US Army TRAC Tactical War Game

Deborah Duong, Christopher Bladon
Agent Based Learning Systems
Birmingham, Alabama, USA
dduong@agentBasedLearningSystems.com

ABSTRACT

Computational social science, as with all complex adaptive systems sciences, involves a great amount of uncertainty on several fronts, including intrinsic arbitrariness due to path dependence, disagreement on social theory and how to capture it in software, input data of different credibility that does not exactly match the requirements of software because it was gathered for another purpose, and inexact matching translations between models that were designed for different purposes than the study at hand. This paper presents a method of formally tracking that uncertainty, keeping the data input parameters proportionate with logical and probabilistic constraints, and capturing proportionate dynamics of the output ordered by the decision points of policy change. Once ordered this way, the data can be compared to other data similarly expressed, whether that data is from simulation excursions or from the real world, for objective comparison and distance scoring at the level of dynamic patterns as opposed to single outcome validation. This method enables wargame adjudicators to be run out with data gleaned from the wargame, enables data to be repurposed for both training and testing set, and facilitates objective validation scoring through soft matching. Artificial intelligence tools used in the method include probabilistic ontologies with crisp and Bayesian inference, game trees that are multiplayer non-zero sum and decision point based rather than turn-based, and Markov processes to represent the dynamic data and align the models for objective comparison.

Probabilistic ontologies are used to model probabilistic matches of variables traded between models, and to model the probabilistic distance between the result
space of the models and real world survey data. These ontologies are used to perform interfacing and validation of the models used to adjudicate the 2010 US Army TRAC Tactical War Game (TWG) - models which simulated Irregular Warfare in Afghanistan. Models include the Cultural Geography population level model and the Nexus Network Learner individual level model. The probabilistic match implements a loose coupling between the multiresolutional and multiperspective models. After many runs, uncertainty of result is expressed in a Markov process to perform validation against time series survey data from Afghanistan, also represented in a Markov process. A distance metric from information theory is used to measure the probabilistic distance between the resultant Markov processes.

**Keywords:** probabilistic ontology, Markov process, agent-based simulation integration, validation, computational experimentation

1 **INTRODUCTION**

This study demonstrates a methodology of Computational Social Science for integrating multiple models and then running validation studies on them. The methodology employs probabilistic ontologies to do the integration and indicate states, and Markov processes to organize the results and perform the validation. The methodology is played out with the models and data of a major study of Irregular Warfare in the US Army - the Training and Doctrine Analysis Center’s Irregular Warfare Tactical War game.

2 **MODEL INTEGRATION USING PROBABILISTIC ONTOLOGIES**

Because the social world is so complex, and because it is so little understood, multiple models and multiple data sources are needed to accurately represent any real-world scenario. Models need to be modular to be switched in and out, so that different plausible theories may be represented such that policies can be tested against them for robustness. Both models and the data to feed those models, in all likelihood, have been created in a different context than for the particular scenario of a study, and will have to be repurposed to fit the needs of the particular combination of models and measurements used in the study. Models and data are thus in multiple perspectives and resolutions, and are in need of translation before they can align correctly. We perform this translation with probabilistic ontologies.

Ontologies represent knowledge in a way that inference may be applied. Ontologies make assertions about domains, or subject matter areas, defining the relationships between the concepts of a domain. Ontologies take the form of taxonomic categories, having more abstract concepts (such as weapon) at higher levels, and more specific concepts (such as AK-47) at lower levels, and having rules
about what makes an object a member of these concepts. Traditional ontologies are “crisp” in that there is no partial membership: either an object belongs to a concept or it doesn’t (it is either an AK-47 or it is not). Aristotelian philosophers of ontology believe this is how it should be, because ontologies are about what exists in a real world regardless of what we know about it. However, we include probability because we choose the complex adaptive systems viewpoint of ontologies, that they should include probabilities, because nature is arbitrary and path-dependent, as well as the cognitive scientist’s point of view of ontologies, that categorization cannot be separated from human cognition, and is by nature uncertain. Computational social science is a field so fraught with uncertainty that relations which might exist and emergent concepts that a computer may data-mine are just as important as definitional concepts.

Ontologies are used to represent data for model integration because they represent the rules that define concepts and can thus translate concepts. Davis noted that semantic interoperation is the biggest problem with integrating models of multiple perspectives; that is, ensuring that model’s meaning of data is consistent with the study’s meaning of the data (Davis and Anderson, 2004). Ontologies are an excellent tool for integration because they ensure consistency by defining the concepts in data, infer relations between data, and flag inconsistencies in data. Each model has its own ontology that defines its concepts, and the particular study has its own ontology that defines its concepts, which the models are supposed to implement. In our design, translations between the concepts of the study and the models occur in a hub and spoke arrangement between the central “hub” ontology of the study and “spoke” model ontologies that implement the study.

3 PROBABILISTIC ONTOLOGY FRAMEWORK

The US Army Training and Doctrine Analysis Center sponsored our analysis of their 2010 Irregular Warfare Tactical Wargame, held in White Sands Missile Range, New Mexico (Schott, 2009). The probabilistic ontologies are implemented with the Probont probabilistic ontology representation, a service of Impact Computing’s open source model integration framework XBM (Makovoz, 2011), combined with the University of Maryland Baltimore County open source BayesOwl probabilistic inference engine (Zhang, 2009).

The author designed Probont specifically for agent based simulation model integration. Probont differs from many probabilistic ontologies in that the probability relations are represented in an OWL ontology (a kind of representation of ontologies for the semantic web) (Duong 2011, pp 67-82). It is important to represent the probabilistic relationships directly in the ontology because they are just as important as the crisp relationships in determining set membership, and when in the same framework, the probabilistic relationships can be used in both sequential and combined inference with the crisp relationships. Probont accomplishes
sequential probabilistic and crisp logic, while BayesOwl accomplishes combined probabilistic and crisp logic that keeps inference consistent. Probont represents probabilistic relationships through a simulation agent paradigm, in the form of Macro-Agents and Micro-Agents. A Macro-Agent is an OWL individual that represents a simulation agent statistically. For example, an agent has a certain chance of having certain demographic characteristics, such as a fifty percent chance having the gender “female, a thirty percent chance that they are the tribe Mongo, given that they inhabit the country Congo. These probabilities are distributions that fill in attributes of the macro agent. The macro agent may have other attributes that are not distributions, and the distributions can depend on variable priors that are not in the same Macro-Agent. The Micro-Agent is an OWL individual generated from the attribute distributions in the Macro-Agent. For example, the Micro-Agent would have the actual attribute of “Female” for Gender and “Mongo” for tribal affiliation. Probont then uses Bayesian inference for multiresolutional model integration, so that a lower level model may have more specific information derived from a higher level model’s statistical trends in a “Macro to Micro conversion,” which is merely the creation of Micro-Agent Individuals from a Macro-Agent individual’s statistics. Conversely, a higher-level model need not receive all the details of what happened to every agent, and may only need to know statistical trends, which are aggregated in the “Micro to Macro conversion.” Both types of inference were used in the analysis of 2010 US Army Irregular Warfare Tactical Wargame.

Because Probont only uses either Bayesian or the “crisp” inference of OWL ontologies, but not both, it is able to do inference in sequence, for example, first Bayesian inference to make Micro-Agent individuals and then crisp inference to categorize them into OWL classes. However, in separating the Bayesian from the traditional inference, we have the possibility of inconsistencies in knowledge. For example, crisp inference with OWL may define that a particular combination of traits in an agent is impossible, but if they are possible in the Macro-Agent’s attribute distributions, then we can’t stop Micro-Agents from being generated with those traits unless we combine logics to prevent inconsistency. Furthermore, combined probabilistic and crisp inference has greater flexibility than inference in sequence, facilitating the combination of statistical knowledge and domain expertise. To combine logics, we borrow the “L-Nodes” from the BayesOwl probabilistic ontology representation. L-nodes, or Logic Nodes, take set theoretic rules such as “and” and “or” rules, and enforce them in the Bayesian inference, so as to constrain the results to follow the rules consistently.

4 PROBABILISTIC ONTOLOGIES OF THE 2010 IRREGULAR WARFARE TACTICAL WARGAME

The “Hub” ontology for the 2010 Irregular Warfare Tactical Wargame includes
the moves that the players can make, and the indicators from the models that are significant to measure in the study (See Figure 1). Our analysis was run to take the models of the wargame, and to run them out using the strategies of the wargame with the help of a game-tree. The hub ontology contains the strategies of the wargame, including decision points, branches, and sequels, as well as goals. It contains all of the rules needed to run the game out (Duong, 2012).

Figure 1. Part of the Hub ontology of the 2010 Irregular Warfare Tactical Wargame, in the OWL ontology browser, Protege.

The ontologies for the individual models are the “spokes” of the study, and for the analysis we choose two of the adjudicating models of the wargame, Nexus and the Cultural Geography model, to be the spokes (see Figure 2).
Each spoke ontology contains the wargame moves that can be entered into the individual models, that aren’t of the same perspective or resolution as that of the Hub. In order to do the translation, a mediation ontology performs the translation between moves. For example, Figure 3 shows how moves of the hub ontology, that the players can make, “CS_CF” (Cordon and Search, Coalition Forces), Provide_Security_CF, and VCP_CF (Vehicle Checkpoint, Coalition Forces) are translated into a move of the Cultural Geography (CG) spoke ontology, “ISAFAttacksTaliban.” The mediation ontology declares, in crisp rules, that in order to translate into the CG move, the hub rule must be accompanied by some events. For example, a cordon and search in the hub becomes a move in which ISAF attacks the Taliban if, during the cordon and search coalition forces find and detain a bomb maker, suicide bomber, or HVI. These rules show up in the CG ontology because all ontologies are simultaneously loaded, and they define the translation to be equivalent to the CG move.
So far, we have gone over the crisp rules of translation. However, the probabilistic rules come in in the creation of the actual move to enter the CG model, in that there is a certain probability that a bomb maker, HVI or suicide bomber would be found on a cordon and search. Actual statistics on these occurrences are entered into the model via distributions of attributes in the Macro Agent. Figure 4 shows the Macro Agent OWL individual “CS_CFMacroInd” that lists a variety of possible events that can occur during a cordon and search, including the ones that translate it into an attack on the Taliban in the CG model (finding and detaining a bombmaker, HVI, or suicide bomber). Each of these OWL properties is filled in with a distribution.

Figure 4. A macro individual for a cordon and search move that contains many distributions for the occurrence of the move’s attributes.

The distribution appears in Figure 5, in a Distribution class of Probont. The distribution class has probability cells, each of which represent a single value in a conditional probability distribution table. The probability cells tell the actual probability (See figure 6). This design is being modified by the authors so as not to create as many individuals that slow down the crisp inference, outside of XBM.
Figure 5. A distribution for a single attribute of the cordon and search, that captures in a conditional probability whether or not a Leader was found and detained.

Figure 6. A single probability cell of the distribution that says there is a 20% chance that a leader was found and detained in a Cordon and Search.

The Micro Agent individual is generated from the probabilistic information in the Macro Agent individual. Figures 7 and 8 show two instances of generation of Micro Agent individuals in Protégé - Figure 8 with an event that causes the OWL crisp inference to classify it as a CG attack on the Taliban, and Figure 7 without such an event. This demonstrates the probabilistic translation of a move, through a sequence of probabilistic and then crisp inference.
Figure 7. The first generated cordon and search does not result in a detained leader, and it is not interpreted as an attack on the Taliban when crisp inference is applied to the individual. Inferred states are in yellow on the left, and do not include an “attack Taliban” move.

Figure 8. The second cordon and search does have a “found and detained leader”, which the rules translate into an ISAFattacksTaliban action in the Cultural Geography model.

5 VALIDATION

The probabilistic ontology further works to keep track of indicator states to measure after many runs that preserve the correct probabilistic relation between the models. The results are then put into a Markov process, to express the probabilistic relation between important indicator states fired by the ontologies, a technique first performed by Bramson (Bramson, 2009). Keeping track of the probabilistic relations between models is a necessary prerequisite to expressing the output in correct proportion. Once the output is expressed in a Markov process, it is compared to another Markov process derived to real world data, with the use of a probabilistic distance function from information theory. This is a validation of the model output at the level of probabilistic patterns rather than single outcomes, necessary because any one outcome is too arbitrary for validation. This method also achieves an objective score, which captures how close the real world outcome is to a typical outcome of the model, as measured by important indicator states. Figure 9
illustrates the comparison.

Figure 9. A comparison of Markov processes of the model and the real world. The normalized distance was 0.21, where 0 is the same Markov process and 1 is the most different Markov process possible for the same indicators.

REFERENCES