Combining Simulation with Machine Learning to Build Accident Models

Robert Alexander and Tim Kelly

Department of Computer Science
University of York, York, YO10 5DD, UK.
{robert.alexander, tim.kelly}@cs.york.ac.uk

Abstract. A crucial part of safety-critical systems development is identifying how system behaviours lead to accidents. Doing this for complex systems is hard because of the need to consider the interaction between multiple simultaneous failures and normal system behaviours. Multi-agent simulation provides a way to explore the behaviour that emerges from a model of a system under normal and failure conditions. However, it is easy to generate vast amounts of data from such simulations that can be hard to comprehend and interpret. Machine learning can be applied to help understand the patterns that are implicit in this data. This paper first describes the concept of hazard analysis, and how this can be performed using multi-agent simulation. It then describes how machine learning techniques can be used to extract rules from simulation output. The approach is illustrated using a military system-of-systems case study.

1 Introduction

A risk-based safety process has hazard analysis at its heart. The *hazards* latent in a safety-critical system are those states of the system that could lead to an accident without any further abnormal occurrences. Once we have a adequately complete set of hazards for a system, the rest of the safety engineering process can proceed from there to prevent their occurrence or to mitigate their effects.

Hazard analysis is the process of identifying hazards present in a system and determining the causal processes that allow them to be reached. A wide range of established hazard analysis techniques exist, but as discussed by the authors in [1] the emerging class of systems referred to as Systems of Systems (SoS) presents many challenges for hazard analysis. There is therefore interest in developing novel hazard analysis approaches. This paper presents such an approach based on the combination of multi-agent simulation and machine learning.

The following section discusses the application of simulation techniques to hazard analysis, and section 3 outlines a systematic method for doing so. Section 4 illustrates the process with a case study based on a military System of Systems. Section 5 discusses the role of the approach in a wider process. Finally, section 6 summarises the paper and describes the expected direction of future work.

2 Hazard Analysis and Simulation

Simulation has been advocated as a tool for safety analysis by a number of authors, including Henk [2] and Johnson [3]. Most existing work, however, concentrates on using stochastic Monte Carlo techniques to acquire quantitative statistical measures of the overall safety of a system under specified conditions. By contrast, the work described in this paper attempts to determine the relationship of simulation parameters to distinct (undesirable) modes of behaviour of the system; the aim is to acquire a useful *qualitative* understanding of system behaviour.

If analysts do want to use these existing tools for hazard analysis, they have to spend considerable time manually studying visualisations of simulation runs. This is unlikely to be practical for performing an adequately thorough analysis.

Analysts therefore need tools that will explicitly explore the space of possible situations, and then clearly show which combinations of conditions will lead to hazards. This paper describes an approach that attempts to achieve this by combining simulation (to explore the behaviour space of the system) and machine learning (to reduce the results to a comprehensible size and complexity).

3 Method

In our approach, an analyst first builds a detailed simulation model (which is simple enough to build, but has dynamics that are too complex to understand just by watching it run), then derives from that various other models that characterise its behaviour (which are simple enough for humans to read and understand). The intention is that these simpler models will guide analysts towards identifying some hazards in the system that they would otherwise have missed.

The simpler models are generated by machine learning algorithms applied the output of the simulation. In order to supply sufficient examples for the learner to work from, a set of deviations is defined on the simulation model and runs are performed for each possible combination of deviations.

One important way to apply deviations to a model, and the one that is used in this paper, is to identify all the external *channels* of each agent over which *interactions* can occur. Examples of channels include network wires, radio transmissions or simply being located in the same airspace. For each channel, a set of *failure modes* of these is then derived (possibly using a set of guide words as in the popular analysis technique HAZOP [4]). Each combination of some entity having some failure mode on some channel provides a single distinct ‘deviation’.

Once the set of deviations has been explored, machine learning techniques are applied to learn a set of rules that relate deviations to accidents. Analysts must study these rules to determine which of them are realistic (as opposed to simulation artifacts) and how the mechanisms of the system gave rise to the accident. From this knowledge the set of hazards present in the system can be derived.

4 Example

The example uses a simulation model of a military unit engaged in anti-guerilla operations. The unit contains Unmanned Air Vehicles (UAVs), artillery pieces and helicopter-borne infantry. A single vignette has been implemented for this model, in which the agents in the system must detect and neutralize a number of static enemy positions.

It is relatively easy to enumerate the possible accidents that can occur in the system. Simple examination of our model reveals that the following accidents are possible:

- Accident 1 — Helicopter collides with another helicopter
- Accident 2 — Helicopter collides with a UAV
- Accident 3 — Landed helicopter is hit by friendly artillery fire
- Accident 4 — UAV collides with a UAV
- Accident 5 — Helicopter hit by enemy anti-aircraft fire

The model was run for a large set of combinations of possible entity-failure pairs, resulting in a large set of logs describing the events that occurred in the simulation runs. Rules were learned from this output using the C4.5 algorithm as described by Quinlan in [5] and implemented in the data mining tool WEKA (described by Witten and Frank in [6]). For example, for the accident 'landed helicopter is hit by artillery fire' the following rules (expressed as entity-failure combinations) were derived:

1. $\neg lossofcommsfailure\_uav2 \ \wedge \ \neg lossofcommsfailure\_uav4 \rightarrow safe$
2. $\neg lossofcommsfailure\_uav2 \ \wedge \ lossofcommsfailure\_uav4 \ \wedge$ $\neg firespreadwidely\_gun3 \ \wedge \ \neg fireskewnorthwest\_gun3 \rightarrow accident$
3. $\neg lossofcommsfailure\_uav2 \ \wedge \ lossofcommsfailure\_uav4 \ \wedge$ $\neg firespreadwidely\_gun3 \wedge fireskewnorthwest\_gun3 \rightarrow safe$
4. $\neg lossofcommsfailure\_uav2 \ \wedge \ lossofcommsfailure\_uav4$ $\wedge \ firespreadwidely\_gun3 \rightarrow accident$
5. $lossofcommsfailure\_uav2 \rightarrow safe$

Table 1 summarises the results of the analysis. For each accident that occurred in the runs performed, it shows the number of instances that contained that accident, the number of rules in the learned model (in the form: total number of rules / number of rules that lead to the accident occurring) and the percentage accuracy of that learned model (over the training set).

It then gives the number of rules above the plausibility threshold (defined to be a probability of 10^{-11} per run) and the highest estimated probability of any of the rules occurring. Here, we have assumed for the sake of illustration an independent probability of 10^{-3} that each failure will be present in a given run. If the data were available, it would be possible to substitute the correct failure rate for each simple system failure. This could be derived by conventional means.

It can be seen from the table that the most likely accidents are the loss of helicopters 2 through 4 to enemy fire, or the collisions between UAVs 1 and 4 or 4 and 3. All the accidents that occurred have at least one rule that can cause them with a probability of 10^{-4} or better. Were this a real system, it is unlikely that this would be considered an acceptable level of safety.

Accident	#runs	#rules	#plausible rules	highest prob	% accuracy
Enemy kills helicopter 1	6657	54/33	19	9.98×10^{-4}	96.7
Enemy kills helicopter 2	6966	42/28	14	1×10^{-3}	99.3
Enemy kills helicopter 3	6738	56/35	21	1×10^{-3}	99.1
Enemy kills helicopter 4	6842	56/35	21	1×10^{-3}	99.2
Artillery kills helicopter 1	14	5/2	2	9.97×10^{-4}	99.5
Artillery kills helicopter 3	14	5/2	2	9.97×10^{-4}	99.5
Collison UAV1 UAV4	2048	2/1	1	1×10^{-3}	100
Collison UAV4 UAV3	3904	2/1	1	1×10^{-3}	100
(other)	0				

Table 1. Summary of the learned rules

For the five accidents identified above, we have rules that correspond to three of them. (Accidents 1 and 2, involving helicopters colliding with UAVs or other helicopters, do not occur in any of the runs we are working with). In section 1 we defined hazards to be “those states of the system that could lead to an accident without any further abnormal occurrences”. These rules therefore describe hazards present in the system.

Some of these hazards may have been predictable by purely manual analysis (e.g. “UAV in shared airspace with no ability to detect other airborne entities”). This is desirable; for the approach to be validated it needs to be able to identify such ‘obvious’ hazards. We expect, however, that it will go further by finding non-obvious hazards. In this example, there are other rules (such as those given above for ‘landed helicopter is hit by artillery fire’) that may not have been predicted, and appear as hazards only through the interaction of the agents in the simulation.

Interpretation of the rules learned has to be part of a process that identifies whether the rules are simulation artifacts, and whether they can be usefully generalised. Full discussion of this is outside the scope of this paper, but the role of simulation in an investigative process is briefly discussed in the following section.

5 Using the Learned Rules

In this work, the aim of the simulation is to identify ways in which hazards (and hence accidents) could reasonably occur; in this respect, it is comparable to existing hazard analysis techniques. Any hazards that are identified through simulation will require further manual investigation — the simulation result is valuable in that it has drawn the analyst’s attention to the hazards and ‘made a case’ for their plausibility by means of the recorded event trace.

Dewar et al, in [7] describe this level of fidelity as ‘weak prediction’. They note that “subjective judgement is unavoidable in assessing credibility” and that when such a simulation produces an unexpected result “it has created an interesting hypothesis that can (and must) be tested by other means”. In other words, when a simulation reveals a plausible system hazard, other, more conventional analyses must be carried out to determine whether it is credible in the real system. Therefore, the role of the simulation analysis is to narrow down a huge analysis space into one that is manually tractable.

6 Summary and Future Work

This paper has described an approach to performing hazard analysis for complex systems of systems using a combination of multi-agent simulation and machine learning. Application to a case study has demonstrated that rules describing hazards can be derived.

Future work will include applying the approach to additional systems and scenarios, and applying other machine learning algorithms. Additional approaches to introducing deviation, along with ways to combine the results from multiple scenarios, will be developed. Finally, in order to evaluate the real-world value of the approach it will be applied in realistic industrial case studies.

7 Acknowledgements

The authors would like to thank Dimitar Kazakov for his assistance in the work described in this paper. The work was funded under the Defence and Aerospace Defence Partnership in High Integrity Real Time Systems (Strand 2).

References

1. Alexander, R., Hall-May, M., Despotou, G., Kelly, T.: Using simulation to evaluate safety policy for systems of systems. In Barley, M., Massacci, F., Mouratidis, H., Unruh, A., eds.: Proceedings of the Second International Workshop on Safety and Security in Multiagent Systems. (2005) 5–21
2. Blom, H.A.P., Stroeve, S.H., de Jong, H.H.: Safety risk assessment by monte carlo simulation of complex safety critical operations. In Redmill, F., Anderson, T., eds.: Proceedings of the Fourteenth Safety-critical Systems Symposium, Bristol, UK, Safety-Critical Systems Club, Springer (2006) 47–67
3. Johnson, C.: The glasgow-hospital evacuation simulator: Using computer simulations to support a risk-based approach to hospital evacuation. Technical report, University of Glasgow (2005) Submitted to the Journal of Risk and Reliability.
4. CISHEC: A Guide to Hazard and Operability Studies. The Chemical Industry Safety and Health Council of the Chemical Industries Association Ltd (1977)
5. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kauffman (1993)
6. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques. 2nd edn. Morgan Kaufmann, San Francisco (2005)
7. Dewar, J.A., Bankes, S.C., Hodges, J.S., Lucas, T., Saunders-Newton, D.K., Vye, P.: Credible uses of the distributed interactive simulation (DIS) system. Technical Report MR-607-A, RAND (1996)