Contingent Scaffolding for System Safety Analysis

Paul S. Brown[1(✉)], Anthony G. Cohn[1], Glen Hart[2], and Vania Dimitrova[1]

[1] University of Leeds, Leeds LS2 9JT, UK
{sc16pb,a.g.cohn,v.g.dimitrova}@leeds.ac.uk
[2] Defence Science and Technology Laboratory (DSTL), Wiltshire SP4 0JQ, UK

Abstract. System safety analysis is a creative process that can often be undertaken by people who are not experts in the system under analysis whilst also learning the analysis methodology. With the increase of system complexity, the high demand for analyses conducted at a scale and the potentially catastrophic consequences of inadequate analysis, there is an urgent need for supporting the development of system analysis skills. Technological solutions can effectively scaffold this ill-defined domain. We propose a generic framework for Contingent Scaffolding capable of providing flexible learning support while conducting system safety analysis. This has been implemented into an intelligent agent, Oswin, which offers **O**ntology-driven **s**caffolding **w**ith **i**nteractive **n**udges.

Keywords: Contingent Scaffolding · System safety · STPA · Ontology

1 Problem Statement

System safety analysis is conducted to understand the behaviour of increasingly complex systems to mitigate or prevent undesirable behaviour. The consequences of inadequate analysis can be catastrophic. To support the analyst several methodologies have been created, one of which is System-Theoretic Process Analysis (STPA) [4]. STPA is relatively new, gaining results comparable with other methodologies and revealing insights they missed [3,8].

Analysts require expert-level knowledge and skills regarding their chosen methodology, chosen model, modelling, as well as the system under consideration. Given that STPA is an emerging methodology, there are a growing number of people wishing to learn it and its associated model. Expertise regarding the system also cannot be assumed given that STPA can be conducted from the design phase, on large systems distributed over teams, and on complex systems requiring expertise in multiple fields.

STPA is an ill-defined task [6] with an ambiguous starting state, an unknown goal state, an advisory non-strict procedure, and no known correct solution. It is an ill-defined domain [6]: STPA is generic to all analyses and thus contains incomplete declarative knowledge regarding a particular analysis, including the system under analysis. System safety is an ill-defined problem [5], in STPA safety

© Springer Nature Switzerland AG 2020
I. I. Bittencourt et al. (Eds.): AIED 2020, LNAI 12164, pp. 395–399, 2020.
https://doi.org/10.1007/978-3-030-52240-7_71

is re-characterised as a control problem, alternative characterisations include Swiss-cheese and dominoes [4].

Contingent Scaffolding is presented by Wood *et al.* [11] as a process enabling the learner to accomplish a task beyond their current capabilities, which is one key goal of supporting the non-expert analyst. It has been successfully applied in Intelligent Tutoring Systems, where it provides graded support for multi-step problem solving in formalised domains [2]. Thus it is used by Oswin as a strategy for delivering feedback as interactive nudges regarding the violation of constraints.

Wood and Wood expounded the principals of "contingent scaffolding" [12] as:

- Help is provided expeditiously when the learner is in trouble
- Help is increased as the learner requires, until the solution is reached
- As the learner succeeds, support is withdrawn

The learner's behaviour is observed to determine whether intervention is required, the tutor then moves through the levels of support. The number of levels vary, between 4 and 5 [12] or 6 [1]; the only guidance being that they should increase in depth or interference until physical intervention is undertaken. There has also been concern in implementations regarding a lack of flexibility [2,10]. It arises from the capability of a learner to approach a problem in an unexpected but valid way. This PhD project takes into account these concerns in the proposed contingency scaffolding framework outlined below. It uses constraints based on situational calculus and a domain ontology to provide scaffolding flexibility in the context of system safety analysis.

2 Proposed Solution and Methodology

Within this project an AI agent, Oswin (**O**ntology-driven **S**caffolding **W**ith **I**nteractive **N**udges), has been prototyped to provide learning support to the non-expert STPA analyst. The intention is to enable them to produce a product beyond their current abilities, whilst improving their knowledge of STPA and the system under analysis, as well as improving their safety-analytic and modelling skills. Oswin uses a constraint-based Contingent Scaffolding framework to accomplish this.

Previous work on ill-defined domains and tasks indicates various strategies have been successful, including constraints [5] which can check if certain properties of a solution are present or not. The violation of some constraint indicates a need for intervention [7]. Oswin is provided constraints as logical-queries over a user-extended ontology, including strong constraints such as a situation can't be both safe and hazardous, as well as advisory constraints such as not analysing more than 7–10 hazards.

The range of help the ontology is capable of supporting exceeds enforcing constraints. It is also capable of providing a reference to factual, conceptual and procedural knowledge as understood by Oswin to ensure a common conceptualisation. Furthermore, it both enables explanations for Oswin's reasoning, and guiding the learner through formulating their own arguments regarding causality or the categorisation of systems. Finally it enables some re-use of systems and their behaviour from previous analyses, encouraging analogous reasoning over multiple analyses: especially beneficial to those specialising in particular system domains such as autonomous vehicles.

The dual issues of flexibility and expeditious intervention are accounted for by the on-line evaluation of constraints [7], and following violations immediately with Contingent Scaffolding. Within this framework, the contingency is formally defined using Situation Calculus. Reiter's definition [9] allows complex reasoning over a log of interactions, including queries over prior situations, which is used to determine fading.

Regarding levels of support for the Contingent Scaffolding Framework, in the absence of specific guidance on the levels to use, successful behaviour of human tutors is used to inform the hierarchy. Due to the nature of the ill-defined domain this hierarchy also accounts for the limitation that it is not always possible to provide a solution as physical intervention. Messages for the first three levels are automatically generated from the constraint, the highest level requires a database of adaptable code snippets that can be executed in the UI to provide physical intervention.

A prototype Oswin has been implemented in Logtalk, based upon the Prolog implementation by Reiter of Situation Calculus [9]. The implementation is split into a Situation Calculus Ontology Authoring tool and a Contingent Scaffolding framework, both of which will be defined in Situation Calculus. The ontology has been defined in Description Logics, OWL, and Prolog. Additional ontological reasoning has been defined in set-builder notation and Prolog.

A prototype interface has also been implemented to facilitate evaluating the efficacy of the provided support, see Fig. 1. It is proposed to test the system on non-expert cohorts who will be provided with STPA training and an example system. Following which they will conduct an analysis independently. Half will have access to Oswin and all will have access to a human with system expertise, simulating an analyst within an organisation. Detailed logs will be gathered via the Situation Calculus implementation, which will then be studied for evaluation.

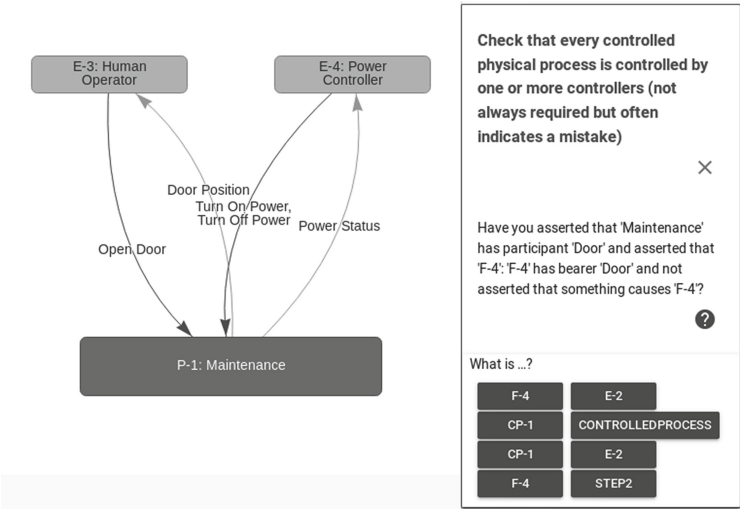

Fig. 1. The UI with Oswin showing level 1 and 2 feedback for a missing "close door" action, which causes "F-4". The user believes they have finished defining all relevant control actions, Oswin believes they have missed one. The ineloquent question asked by Oswin is generated from and reflects the successful unification of the constraint query used to arrive at its belief. The available interactions Oswin provides are to request more help, dismiss the nudge (Oswin may be wrong), or to lookup a relevant term. By defining the missing control action, the feedback nudge will dismiss itself with no direct interaction.

3 Expected Contributions and Future Work

The main contribution of this PhD project to AIED is a framework for ontology-driven scaffolding with interactive nudges for developing system safety analysis skills. It uses situational calculus and a domain ontology to specify situations requiring scaffolding and to automatically generate interactive nudges. While the framework is illustrated in system safety analysis, providing a formal, logical specification enables generalisation to similar ill-defined domains and tasks (e.g. debugging, software security, design).

Currently we have a working prototype of Oswin, using the framework in the system safety domain. Preliminary testing has been conducted with a representative STPA example (interlock system [4]) by a small group of system safety analysis novices. Our immediate work is an evaluation of the efficacy of the application in the challenging domain provided by STPA. It is expected that Oswin users' final ontological models representing the outcome of the system safety analysis will be close to expert ones. Additional analysis will consider non-productive behaviour, timings, and resolution of interventions. We also consider retention of learning and re-use of system safety analysis patterns and components across different scenarios.

Acknowledgements. The authors gratefully acknowledge the financial support provided: an EPSRC CASE studentship partially funded by the Defence Science and Technology Laboratory (Dstl). The advice provided by experts at Dstl is also acknowledged.

References

1. Daniels, H.: Vygotsky and Pedagogy. Routledge, London (2010)
2. Du Boulay, B., Luckin, R.: Modelling human teaching tactics and strategies for tutoring systems. Int. J. Artif. Intell. Educ. **12**, 235–256 (2001)
3. Fleming, C.H., Spencer, M., Thomas, J., Leveson, N., Wilkinson, C.: Safety assurance in NextGen and complex transportation systems. Saf. Sci. **55**, 173–187 (2013)
4. Leveson, N.: Engineering a Safer World. The MIT Press, Cambridge (2017)
5. Lynch, C., Ashley, K.D., Pinkwart, N., Aleven, V.: Concepts, structures, and goals: Redefining ill-definedness. Int. J. AI Educ. **19**(3), 253–266 (2009)
6. Mitrovic, A., Weerasinghe, A.: Revisiting ill-definedness and the consequences for ITSs. In: Artificial Intelligence in Education: Building Learning Systems that Care from Knowledge Representation to Affective Modelling, pp. 375–382 (2009)
7. Ohlsson, S.: Constraint-based modeling: From cognitive theory to computer tutoring - and back again. Int. J. Artif. Intell. Educ. **26**(1), 457–473 (2015)
8. Pawlicki, T., Samost, A., Brown, D.W., Manger, R.P., Kim, G.Y., Leveson, N.G.: Application of systems and control theory-based hazard analysis to radiation oncology. Med. Phys. **43**(3), 1514–1530 (2016)
9. Reiter, R.: Knowledge in Action. The MIT Press, Cambridge (2001)
10. Wood, D.: Commentary: Contribution of scaffolding to learning and teaching: Interdisciplinary perspectives. Int. J. Educ. Res. **90**, 248–251 (2018)
11. Wood, D., Bruner, J.S., Ross, G.: The role of tutoring in problem solving. J. Child Psychol. Psychiatry **17**(2), 89–100 (1976)
12. Wood, H., Wood, D.: Help seeking, learning and contingent tutoring. Comput. Educ. **33**(2–3), 153–169 (1999)