FRAMEWORK FOR PROBABILISTIC ANALYSIS OF PROGRAMS AND LINKING TO SCIENTIFIC LITERATURE

by

Paul Douglas Hein

A Thesis Submitted to the Faculty of the

DEPARTMENT OF COMPUTER SCIENCE

In Partial Fulfillment of the Requirements $\mbox{For the Degree of}$

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

THE UNIVERSITY OF ARIZONA GRADUATE COLLEGE

As members of the Master's Committee, we certify that we have read the thesis prepared by Paul Douglas Hein, titled Framework for Probabilistic Analysis of Programs and Linking to Scientific Literature and recommend that it be accepted as fulfilling the thesis requirement for the Master's Degree.

	Date: 3 May 2019
Clayton Morrison	-
	Date: 3 May 2019
Mihai Surdeanu	-
	Date: 3 May 2019
Saumya Debray	
Final approval and acceptance of this thesis is conting submission of the final copies of the thesis to the Gradua	-
I hereby certify that I have read this thesis prepared un commend that it be accepted as fulfilling the Master's req	· ·
	Date: 3 May 2019
Clayton Morrison	
Associate Professor	
School of Information	

TABLE OF CONTENTS

LIST OF FIGURES	4
LIST OF TABLES	1
ABSTRACT	6
CHAPTER 1 INTRODUCTION	7
CHAPTER 2 SOURCE CODE MAPPING	8
CHAPTER 3 VARIABLE GROUNDING	Ć
4.2 Common Subnetwork Isolation	1() 1() 1()
5.1 Sampling Techniques	11 11 11 12
CHAPTER 6 SOURCE CODE SUMMARIZATION	13
APPENDIX A Sample Appendix	14
REFERENCES	l :

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

INTRODUCTION

Some intro.

1.1 Overview

Some text.

SOURCE CODE MAPPING

Some intro about the task of mapping a program in Fortran (or some other language) to our GrFN interlingua.

2.1 Function Extraction

Some text about extracting lambda functions and a GrFN spec from source code.

2.2 Grounded Function Network Generation

Some text about creating GrFNs.

VARIABLE GROUNDING

Some intro about the task of grounding variables

AUTOMATED MODEL COMPARISON

Some intro about the task of comparing two or more GrFN models that have been extracted from source code.

4.1 Variable Node Identification

Some text about the task of unifying variable nodes that may have different names but represent the same physical variable.

4.2 Common Subnetwork Isolation

Some text about identifying the Forward Influence Blanket of two or more models that isolates the shared components of two models in a markov blanket that allows for forward inference.

4.3 Unrelated Subgraph Diff

Some text that discuss the challenges and our solutions for comparing the subcomponents of two GrFNs that do not share any information in common.

AUTOMATED MODEL ANALYSIS

Some intro about the task of analyzing two or more GrFN models that have been extracted from source code, or the Forward Influence Blanket of two or more GrFNs.

5.1 Sampling Techniques

Some text about the basic method for sampling from the inputs to a GrFN given little information.

5.1.1 Saltelli Sampling

This section should introduce Saltelli sampling and go into great detail on the process.

5.1.2 Data Informed Sampling Techniques

This section should introduce the methods that we can use to sample from preexisting data files associated with the codebase that we are working with.

5.1.3 Grounded Sampling Techniques

This section should go into detail on how we can use the grounded nature of our variables to get information about how they exists in the real world that we can use to constrain the domains for our inputs.

5.2 Sensitivity Analysis

Some text that discusses our methods for conducting Sensitivity analysis of our extracted GrFNs.

5.2.1 Variance Based Analysis

This section will go in depth on the Sobol method of Sensitivity analysis.

5.3 Output Surface Analysis

This section will introduce the idea of generating output surfaces that scientists can view and interact with in order to better understand the sensitivity of their models based upon key inputs. The choice in which surfaces to show will be directed by sensitivity analysis.

SOURCE CODE SUMMARIZATION

Some intro about the task summarizing source code

6.1 Embedding Source Code Tokens

Some text about creating a good embedding scheme for source code

6.2 Seq-2-Seq Documentation Generation

Some text about translating a documentation sequence from a source code sequence

APPENDIX A

Sample Appendix

Stuff.....

REFERENCES