Ranjeev Mittu · Donald Sofge
Alan Wagner · W.F. Lawless *Editors*

Robust Intelligence and Trust in Autonomous Systems

Springer

Robust Intelligence and Trust in Autonomous Systems

Ranjeev Mittu • Donald Sofge • Alan Wagner
W.F. Lawless
Editors

Robust Intelligence and Trust in Autonomous Systems

Springer

Editors
Ranjeev Mittu
Naval Research Laboratory
Washington, DC, USA

Donald Sofge
Naval Research Laboratory
Washington, DC, USA

Alan Wagner
Georgia Tech Research Institute
Atlanta, GA, USA

W.F. Lawless
Paine College
Augusta, GA, USA

ISBN 978-1-4899-7666-6 ISBN 978-1-4899-7668-0 (eBook)
DOI 10.1007/978-1-4899-7668-0

Library of Congress Control Number: 2015956336

Springer New York Heidelberg Dordrecht London

Printed on acid-free paper

Springer Science+Business Media LLC New York is part of Springer Science+Business Media (www.springer.com)

Preface

This book is based on the Association for the Advancement of Artificial Intelligence (AAAI) Symposium on "The Intersection of Robust Intelligence (RI) and Trust in Autonomous Systems"; the symposium was held at Stanford March 24–26, 2014. The title of this book reflects the theme of the symposium. Our goal for this book is to further address the current state of the art in autonomy at the intersection of RI and trust and to more fully examine the existing research gaps that must be closed to enable the effective integration of autonomous and human systems. This research is particularly necessary for the next generation of systems, which must scale to teams of autonomous platforms to better support their human operators and decision makers.

The book explores the intersection of RI and trust across multiple contexts and among arbitrary combinations of humans, machines, and robots. To help readers better understand the relationships between artificial intelligence (AI) and RI in a way that promotes trust among autonomous systems and human users, this edited volume presents a selection of the underlying theories, computational models, experimental methods, and possible field applications. While other books deal with these topics individually, this book is unique in that it unifies the fields of RI and trust and frames them in the broader context of effective integration for human-autonomous systems.

The volume begins by describing the current state of the art for research in RI and trust presented at Stanford University in the Spring of 2014 (copies of the technical articles are available from AAAI at http://www.aaai.org/Library/Symposia/Spring/ss14-04.php; a link to the presentation materials and photographs of participants is at https://sites.google.com/site/aaairobustintelligence/).

After the introduction, chapter contributors elaborate on key research topics at the heart of effective human-systems integration. These include machine learning, Big Data, workload management, human-computer interfaces, team integration and performance, advanced analytics, behavior modeling, training, and test and evaluation, the latter known as V&V (i.e., verification and validation).

The contributions to this volume are written by world-class leaders from across the field of autonomous systems research, ranging from industry to academia and to government. Given the diversity of the research in this book, we strove to thoroughly examine the challenges and trends of systems that exhibit RI; the fundamental implications of RI in developing trusted relationships among humans, machines, and robots with present and future autonomous systems; and the effective human systems integration that must result for trust to be sustained.

A brief summary is presented below of the AAAI Symposium in the Spring of 2014.

AAAI-2014 Spring Symposium Organizers

Jennifer Burke, Boeing: jennifer.l.burke2@boeing.com
Alan Wagner, Georgia Tech Research Institute: Alan.Wagner@gtri.gatech.edu
Donald Sofge, Naval Research Laboratory: don.sofge@nrl.navy.mil
William F. Lawless, Paine College: wlawless@paine.edu

AAAI-2014 Spring Symposium: Keynote Speakers

- Suzanne Barber, barber@mail.utexas.edu, AT&T Foundation Endowed Professor in Engineering, Department of Electrical and Computer Engineering, Cockrell School of Engineering, U Texas
- Julie L. Marble, julie.marble@navy.mil, Program Officer: Hybrid human computer systems at Office of Naval Research, Washington, DC
- Ranjeev Mittu, ranjeev.mittu@nrl.navy.mil, Branch Head, Information Management & Decision Architectures Branch, Information Technology Division, US Naval Research Laboratory, Washington, DC
- Hadas Kress-Gazit, hadaskg@cornell.edu, Cornell University; High-Level Verifiable Robotics
- Satyandra K. Gupta, skgupta@umd.edu, Director, Maryland Robotics Center, University of Maryland
- Dave Ferguson, daveferguson@google.com, Google's Self-Driving Car project, San Francisco
- Mo Jamshidi, mo.jamshidi@usta.edu, University of Texas at San Antonio, Lutcher Brown Endowed Chair and Professor, Computer and Electrical Engineering
- Dirk Helbing, dirk.helbing@gess.ethz.ch, http://www.futurict.eu; ETH Zurich

Symposium Program Committee

- Julie L. Marble, julie.Marble@jhuapl.edu, cybersecurity, Johns Hopkins Advanced Physics Lab, MD
- Ranjeev Mittu, ranjeev.mittu@nrl.navy.mil, Branch Head, Information Management & Decision Architectures Branch, Information Technology Division, U.S. Naval Research Laboratory, Washington, DC
- David Atkinson, datkinson@ihmc.us, Senior Research Scientist, Institute of Human-Machine Cognition (IHMC)
- Jeffrey Bradshaw, jbradshaw@ihmc.us; Senior Research Scientist, Institute of Human-Machine Cognition (IHMC)
- Lashon B. Booker, booker@mitre.org, The MITRE Corporation
- Paul Hyden, paul.hyden@nrl.navy.mil, Naval Research Laboratory
- Holly Yanco, holly@cs.uml.edu, University of Massachusetts Lowell
- Fei Gao, feigao@MIT.EDU.MIT
- Robert Hoffman, rhoffman@ihmc.us, Senior Research Scientist, Institute of Human-Machine Cognition (IHMC)
- Florian Jentsch, florian.Jentsch@ucf.edu, Department of Psychology and Institute for Simulation & Training, *Director*, Team Performance Laboratory, University of Central Florida
- Howell, Chuck, howell@mitre.org, Chief Engineer, Intelligence Portfolio, National Security Center, The MITRE Corporation
- Paul Robinette, probinette3@gatech.edu, Graduate Research Assistant, Georgia Institute of Technology
- Munjal Desai, munjaldesai@google.com
- Geert-Jan Kruijff, gj@dfki.de, Senior Researcher/Project Leader, Language Technology Lab, DFKI GmbH, Saarbruecken, Germany

This AAAI symposium sought to address these topics and questions:

- How can robust intelligence be instantiated?
- What is RI for an individual agent? A team? Firm? System?
- What is a robust team?
- What is the association between RI and autonomy?
- What metrics exist for robust intelligence, trust, or autonomy between individuals or groups, and how well do these translate to interactions between humans and autonomous machines?
- What are the connotations of "trust" in various settings and contexts?
- How do concepts of trust between humans collaborating on a task differ from human-human, human-machine, machine-human, and machine-machine trust relationships?
- What metrics for trust currently exist for evaluating machines (possibly including such factors as reliability, repeatability, intent, and susceptibility to catastrophic failure), and how may these metrics be used to moderate behavior in collaborative teams including both humans and autonomous machines?

- How do trust relationships affect the social dynamics of human teams, and are these effects quantifiable?
- What validation procedures could be used to engender trust between a human and an autonomous machine?
- What algorithms or techniques are available to allow machines to develop trust in a human operator or another autonomous machine?
- How valid are the present conceptual models of human networks? Mathematical models? Computational models?
- How valid are the present conceptual models of autonomy in networks? Mathematical models? Computational models?

Papers at the symposium specified the relevance of their topic to AI or proposed a method involving AI to help address their particular issue. Potential topics included (but were not limited to) the following:

Robust Intelligence (RI) topics:

- Computational, mathematical, conceptual models of robust intelligence
- Metrics of robust intelligence
- Is a model of thermodynamics possible for RI (i.e., using physical thermodynamic principles, can intelligent behavior be addressed in reaction to thermodynamic pressure from the environment?)?

Trust topics:

- Computational, mathematical, conceptual models of trust in autonomous systems
- Human requirements for trust and trust in machines
- Machine requirements for trust and trust in humans
- Methods for engendering and measuring trust among humans and machines
- Metrics for deception among humans and machines
- Other computational and heuristic models of trust relationships, and related behaviors, in teams of humans and machines

Autonomy topics:

- Models of individual, group, and firm autonomous system behaviors
- Mathematical models of multitasking in a team (e.g., entropy levels overall and by individual agents, energy levels overall and by individual agents)

Network topics:

- Constructing, measuring, and assessing networks (e.g., the density of chat networks among human operators controlling multi-unmanned aerial vehicles)
- For networks, specify whether the application is for humans, machines, robots, or a combination, e.g., the density of inter-robot communications

After the symposium was completed, the book and the symposium took on separate lives. The following individuals were responsible for the proposal submitted to Springer after the symposium, for the divergence between the topics of the two, and for editing the book that has resulted.

Washington, DC, USA Ranjeev Mittu
Washington, DC, USA Donald Sofge
Atlanta, GA, USA Alan Wagner
Augusta, GA, USA W.F. Lawless

Contents

Chapter 1
Introduction

RanjeevMittu, Donald Sofge, AlanWagner, and W. F. Lawless

1.1 The Intersection of Robust Intelligence (RI) and Trust in Autonomous Systems

The Intersection of Robust Intelligence (RI) and Trust in Autonomous Systems addresses the current state-of-the-art in autonomy at the intersection of Robust Intelligence (RI) and trust, and the research gaps that must be overcome to enable the effective integration of autonomous and human systems. This is particularly true for the next generation of systems, which must scale to teams of autonomous platforms to better support their human operators and decision makers. This edited volume explores the intersection of RI and trust across multiple contexts among autonomous hybrid systems (where hybrids are arbitrary combinations of humans, machines and robots). To better understand the relationships between Artificial Intelligence (AI) and RI in a way that promotes trust between autonomous systems and human users, this edited volume explores the underlying theory, mathematics, computational models, and field applications.

To better understand and manage RI with AI in a manner that promotes trust in autonomous agents and teams, our interest is in the further development of theory, network models, mathematics, computational models, associations, and field

R. Mittu • D. Sofge
Naval Research Laboratory, 4555 Overlook Ave SW, Washington, DC 20375, USA
e-mail: ranjeev.mittu@nrl.navy.mil; donald.sofge@nrl.navy.mil

A. Wagner
Georgia Tech Research Institute, 250 14th Street NW, Atlanta, GA 30318, USA
e-mail: Alan.Wagner@gtri.gatech.edu

W.F. Lawless (✉)
Paine College, 1235 15th Street, Augusta, GA 30901, USA
e-mail: WLawless@paine.edu

R. Mittu et al. (eds.), *Robust Intelligence and Trust in Autonomous Systems*,
DOI 10.1007/978-1-4899-7668-0_1

applications at the intersection of RI and trust. We are interested not only in effectiveness with a team's multitasking or in constructing RI networks and models, but in the efficiency and trust engendered among interacting participants.

Part of our symposium in 2014 sought a better understanding of the intersection of RI and trust for humans interacting with other humans and human groups (e.g., teams, firms, systems; also, the networks among these social objects). Our goal is to use this information with AI to not only model RI and trust, but also to predict outcomes from interactions between autonomous hybrid groups (e.g., hybrid teams in multitasking operations).

Systems that learn, adapt, and apply their experience to the problems faced in an environment may be better suited to respond to new and unexpected challenges. One could argue that such systems are "robust" to the prospect of a dynamic and occasionally unpredictable world. We expect the systems that exhibit this type of robustness to afford to those who interact with the system a greater degree of trust. For instance, an autonomous vehicle which, in addition to driving to different locations by itself, can also warn a passenger of locations where it should not drive, might likely be viewed as more robust than a similar system without such a warning capability. But would it be viewed as more trustworthy? This workshop endeavored to examine such questions that lay at the intersection of robust intelligence and trust. Problems such as these are particularly difficult because they imply situational variations that may be hard to define.

The focus of our workshop centered on how robust intelligence impacts trust in the system and how trust in the system makes it more or less robust. We explored approaches to RI and trust that included, among others, intelligent networks, intelligent agents, and multitasking by hybrid groups (i.e., arbitrary combinations of humans, machines and robots).

1.2 Background of the 2014 Symposium

Robust intelligence (RI) has not been easy to define. We proposed an approach to RI with artificial intelligence (AI) that may include, among other approaches, the science of intelligent networks, the generation of trust among intelligent agents, and multitasking among hybrid groups (humans, machines and robots). RI is the goal of several government projects to explore the intelligence as seen at the level of humans, including those directed by NSF (2013); the US Army (Army 2014) and the USAF (Gluck 2013). DARPA (2014) has a program on physical intelligence that is attempting to produce the first example of ""intelligent" behavior under thermodynamic pressure from their environment." Carnegie Mellon University (CMU 2014) has a program to build a robot that can execute "complex tasks in dangerous … environments." IEEE (2014) has the journal *Intelligent Systems* to address various topics on intelligence in automation including trust; social computing; health; and, among others, coalitions that make the "effective use of limited resources to achieve complex and multiple objectives." From another