Lecture Notes in Artificial Intelligence 1209

Subseries of Lecture Notes in Computer Science
Edited by J. G. Carbonell and J. Siekmann

Lecture Notes in Computer Science
Edited by G. Goos, J. Hartmanis and J. van Leeuwen

Lawrence Cavedon Anand Rao
Wayne Wobcke (Eds.)

Intelligent Agent Systems

Theoretical and Practical Issues

Based on a Workshop Held at PRICAI '96
Cairns, Australia, August 26-30, 1996

Springer

Series Editors
Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Lawrence Cavedon
Royal Melbourne Institute of Technology, Computer Science Department
124 La Trobe Street, Melbourne, Victoria 3000, Australia
E-mail: cavedon@cs.rmit.edu.au

Anand Rao
Australian Artificial Intelligence Institute
Level 6, 171 La Trobe Street, Melbourne, Victoria 3000, Australia
E-mail: anand@aaii.oz.au

Wayne Wobcke
University of Sydney, Basser Department of Computer Science
Sydney, NSW 2006, Australia
E-mail: wobcke@cs.su.oz.au

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Intelligent agent systems : theoretical and practical issues ; based on a workshop held at PRICAI '96, Cairns, Australia, August 26 - 30, 1996. Lawrence Cavedon ... (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Budapest ; Hong Kong ; London ; Milan ; Paris ; Santa Clara ; Singapore ; Tokyo : Springer, 1997
(Lecture notes in computer science ; Vol. 1209 : Lecture notes in artificial intelligence)
ISBN 3-540-62686-7
NE: Cavedon, Lawrence [Hrsg.]; PRICAI <4, 1996, Cairns>; GT

CR Subject Classification (1991): I.2, D.2, C.2.4

ISBN 3-540-62686-7 Springer-Verlag Berlin Heidelberg New York

Typesetting: Camera ready by author
SPIN 10550374 06/3142 - 5 4 3 2 1 0 Printed on acid-free paper

Preface

This volume emanated from a workshop on the Theoretical and Practical Foundations of Intelligent Agents, held at the Fourth Pacific Rim International Conference on Artificial Intelligence in Cairns, Australia, in August, 1996. The aim of the workshop was to bring together researchers working on formal aspects of rational agency, novel agent architectures, and principles underlying the implementation of agent-based systems.

The papers presented at the workshop were revised (in some cases quite substantially) following comments from referees and workshop participants. Further papers were solicited from a number of leading researchers in the area who also served on the programme committee of the workshop: John Bell, Yves Lesperance, Jörg Müller, and Munindar Singh.

The papers in this volume have been grouped around three main topics: agent architectures, formal theories of rationality, and cooperation and collaboration. The papers themselves represented a broad cross-section of topics within these areas, including software agents, BDI architectures,[1] social commitment, believable agents, and Artificial Life. The workshop also included an invited presentation by Professor Rodney Brooks of Massachusetts Institute of Technology, and a panel discussion on the role of beliefs, desires, and intentions in the design of autonomous agents (not reproduced in this volume).

Agent architectures

This section contains an extensive survey of control architectures for autonomous agent systems (Müller) and papers covering issues related to agent architectures, including software agents (Cranefield and Purvis) and believable agents (Padgham and Taylor). Such issues are gathering increasing importance, particularly with the growing interest in distributed applications and animated interfaces. Also included is a paper on the use of a logic programming language (Golog) to build agent systems (Lesperance et al.). Golog represents a significant link between formal theory (for reasoning about action) and practice.

Müller's article surveys various architectures proposed for reactive agents, deliberative agents, interacting agents, hybrid approaches (i.e., combining the reactive and deliberative approaches), as well as believable agents (i.e., agents with "personality"), software agents, and softbots. This article is in the same style as (but significantly extends) part of Wooldridge and Jennings's earlier survey article.[2]

Lesperance, Levesque, and Ruman describe the use of the logic programming language Golog in building a software agent system to support personal banking

[1] A number of papers in this volume are based on a BDI framework, in which rational behavior is described in terms of an agent's *beliefs, desires,* and *intentions,* drawing particularly on the work of: Bratman, M.E. *Intentions, Plans and Practical Reason*, Harvard University Press, Cambridge, MA, 1987.

[2] Wooldridge, M. and Jennings, N.R. "Intelligent agents: theory and practice," *Knowledge Engineering Review,* **10**, 1995.

over networks. Golog is based on a situation calculus approach to reasoning about actions and their effects. Programming in Golog involves writing out the preconditions and expected effects of actions as axioms. The Golog interpreter constructs proofs from these axioms, which enable the prediction of the effects of various courses of action; these are then used to select the most appropriate action to perform.

Cranefield and Purvis present a software agent approach to coordinating different software tools to perform complex tasks in a specified domain. Their architecture combines a planning agent with KQML-communicating agents that control the particular software tools. Goals (tasks requiring multiple subtasks) are sent to the planner, which plans a sequence of actions, each of which is performed by some tool agent. New tool agents are incorporated into the architecture by specifying the preconditions and effects of each action that can be performed in terms of the abstract view of the data on which the actions operate.

Padgham and Taylor describe an architecture for believable agents formed by extending a BDI agent architecture with a simple model of emotions and personality. In their model, the way an agent reacts to goal successes and failures affects both the agent's choice of further actions and the depiction of that agent as an animated figure. Personality is a function of the emotions and "motivational concerns" of an agent, as well as how predisposed the agent is in reacting to those emotions.

Formal theories of rationality

The papers in this section all concern the formal foundations of agents. Two papers address the formalization of the notion of commitment: in particular, the relationship of commitments to resource-boundedness (Singh) and to goal maintenance and coherence (Bell and Huang). A further two papers address issues related to formalizing "practical" as opposed to "idealized" rational agents: one paper presents a model of a "limited reasoner" (Moreno and Sales) while the other provides a model of a particular implemented system (Morley). The final paper in this section (van der Meyden) shows some interesting connections between logical specifications of knowledge-based protocols and their implementations.[3] These last three papers reflect a current trend towards research bridging the gap between theory and practice.

Singh investigates the notion of commitment within a BDI framework. After reviewing the way in which such commitment is of benefit to resource-bounded agents, he discusses and formalizes an approach to *precommitments*. A precommitment is effectively a long-term commitment to an intention, one that is not normally reconsidered during the agent's deliberative reasoning. Adopting a precommitment has several consequences for an agent, e.g., the cost of satisfying

[3] This is related to the work of: Rosenschein, S.J. and Kaelbling, L.P. "The synthesis of digital machines with provable epistemic properties," in Halpern, J.Y. (Ed.) *Theoretical Aspects of Reasoning About Knowledge: Proceedings of the 1986 Conference*, Morgan Kaufmann, Los Altos, CA, 1986.

the corresponding commitment may be increased, or the option of adopting that commitment (in the future) may be ruled out altogether. The potential benefit to the agent is that deliberation may be less resource-intensive after the precommitment has been made.

Bell and Huang also consider the concept of strength of commitment. They present a logical framework under which an agent's goals are arranged in a hierarchy depending on the strength of the agent's commitment to them. Within this framework, Bell and Huang address issues such as the mutual coherence of an agent's goals, and the revision of goals when multiple goals are not simultaneously achievable.

Moreno and Sales present a model of a "limited reasoner" within an approach to agent design based on a syntactic view of possible worlds. An agent's inference method is implemented using a semantic tableau proof procedure; imperfect reasoning arises from limiting the class of rules used in proof construction. Morley uses a logic of events to provide a detailed formal model of a particular BDI agent architecture (Georgeff and Lansky's *PRS*[4]). Morley's logic of actions and events is specifically designed to handle parallel and composite events, making it particularly suited to modeling multi-agent and dynamic environments.

Van der Meyden investigates the automatic generation of finite state implementations of knowledge-based programs, i.e., programs whose action-selection is specified in terms of their "knowledge" (represented using modal logic) of their environment.[5] In particular, he defines a sufficient condition under which a finite state implementation of a knowledge-based program exists, under the assumption of *perfect recall*—the assumption that a program has full knowledge of its previous observations—and defines a procedure (using concepts from the formal analysis of distributed systems) that constructs an efficient finite-state implementation for such programs.

Cooperation and collaboration

The papers in this section address the extension of models of single agents to multiple cooperating or competing agents. The first paper addresses the formalization of commitment between agents in a multi-agent setting (Cavedon et al.). The remaining papers adopt an experimental methodology: one paper presents some interesting effects in a multi-agent version of the Tileworld that arise from simplified communication between agents (Clark et al.); the second presents an approach to the prisoner's dilemma in an Artificial Life environment (Ito).

Cavedon, Rao and Tidhar investigate the notion of *social commitment*: i.e., commitment between agents, as opposed to the "internal" commitment of an agent to a goal (e.g. as in Singh's paper). Cavedon et al. describe preliminary

[4] Georgeff, M.P. and Lansky, A.L. "Reactive reasoning and planning," *Proceedings of the Sixth National Conference on Artificial Intelligence*, 1987.

[5] An extensive introduction to this approach is given in: Fagin, R., Halpern, J.Y., Moses, Y. and Vardi, M.Y. *Reasoning About Knowledge*, MIT Press, Cambridge, MA, 1995.

work towards formalizing Castelfranchi's notion of social commitment[6] within a BDI logic. They relate the social commitments of a team to the internal commitments of the agents (though not totally reducing the former to the latter), and formally characterize a variety of social behaviors an agent may display within a team environment.

Clark, Irwig, and Wobcke describe a number of experiments using simple BDI agents in the Tileworld, a simple dynamic testbed environment, investigating the "emergence" of benefits to agents arising from simple communication of an agent's intentions to other nearby agents. Their most interesting result is that under certain circumstances, the performance of individual agents actually improves as the number of agents increases, despite a consequent increase in competition for resources due to those agents.

Ito adopts an Artificial Life approach to the iterated Prisoner's Dilemma game, and shows that the "dilemma" can be overcome in a setting involving disclosure of information: i.e., in which agents are required to honestly disclose their past social behavior. The experimental results show that a population of cooperative agents eventually dominates a population of more selfish agents under these conditions.

Acknowledgements

The editors would like to thank the organizers of the Fourth Pacific Rim International Conference on Artificial Intelligence (PRICAI'96) for their support of the workshop, and all the workshop participants. We would particularly like to thank the programme committee members. Cavedon and Rao were partially supported by the Cooperative Research Centre for Intelligent Decision Systems.

Programme Committee

John Bell	Queen Mary and Westfield College, UK
David Israel	SRI International, USA
Yves Lesperance	York University, Canada
Jörg Müller	Mitsubishi Electric Digital Library Group, UK
Ei-Ichi Osawa	Computer Science Laboratory, Sony, Japan
Munindar Singh	North Carolina State University, USA
Liz Sonenberg	University of Melbourne, Australia

Workshop Organizers

Lawrence Cavedon	Royal Melbourne Institute of Technology, Australia
Anand Rao	Australian Artificial Intelligence Institute, Australia
Wayne Wobcke	University of Sydney, Australia

[6] Castelfranchi, C. "Commitments: from individual intentions to groups and organizations," *Proceedings of the First International Conference on Multi-Agent Systems*, 1995.

Contents