A Causal Framework for Intelligent Agents and Multiagent Systems

Hector G. Ceballos

ceballos@itesm.mx
Tecnologico de Monterrey
Av. Eugenio Garza Sada 2501, Monterrey, N.L. Mexico

1 Motivation

Automation of organizational processes that make intensive use of expert knowledge has on the intelligent agent paradigm an appropriate metaphor. The idea would be introducing autonomous agents carrying out tasks on behalf of human users in organizational processes.

Knowledge transference from experts to autonomous agents has been treated on different ways taking advantage of more than 20 years of experience in knowledge representation in Artificial Intelligence. Data mining and case-based reasoning use historical information for identifying patterns that guide agent's decision. Inference rules are used for explicitly capturing expert knowledge. Agent's capacity for adapting to changes rely on the employed formalism. In many cases, the expert is left aside in order to provide an automatic solution.

On the other hand, Autonomic Computing is providing interesting approaches for dealing with the dynamic adaptation of components to organizational and environmental changes. This theory proposes incorporating self-configuration, self-optimization, self-healing and self-protecting capabilities on information systems. Solutions on this field are usually partial and seem to lack a coherent framework for representing the notion of integral autonomy.

In order to propose a conceptual framework that provide coherence on the modeling of autonomic systems we are introducing some notions of Classical Philosophy that have not been taken full advantage of. This framework intends to facilitate knowledge transference from human experts to autonomous agents and, on the opposite direction, the participation of human users on the adaptation to environmental changes.

2 Philosophical foundations

Classical Causality theory [1] refers to necessary relationship between one event (cause) and another event (effect) that is the direct consequence of the first. Aristotle categorizes all causes in four main types: formal cause (the form issued), material cause (materials transformed in the process), efficient cause (the entity responsible by change), and final cause (the intention of the efficient cause).

Forms are explained in metaphysics through the notion of *essence*. An essence is described in terms of potential properties that all the individuals issuing such form might show. Whereas matter is actuality, form is potentiality. Aristotle classifies all predicates that can be said about an individual (in act or in potency) in ten categories: essence, quantitative, qualitative, relational, time, place, possession, disposition, action and affection.

3 Hypothesis

The main hypothesis of this work is: "*Incorporating notions of classical metaphysics and causality on the design of multi-agent systems can improve the relation between human and software agents.*"

Notions borrowed from classical philosophy are represented by four main principles that constitute what we propose to call *Causal Artificial Intelligence Design* (CAID) theory:

I. Total Causality. Any software agent or system can be specified and instantiated through the four main causes: formal, material, efficient and final.
II. Intentional causation. Every component of a system is created with a purpose (final cause) derived from an upper level plan.
III. Ontological commitment. Agents will commit to achieve only goals consistent with its essence (formal cause) and will act in consequence.
IV. Causal effect accountability. Agent decisions are based on the accountable causal effects of its actions.

4 Methodology

To determine if CAID principles are sufficient for developing agents capable of adapting dynamically to institutional and environmental changes the following methodology is being used:

1. Use of Description Logics formalisms for representing agent classes and expert knowledge. Agent classes are used for describing human and software agents in terms of their potential properties, including actions. Conjunctive queries are used for representing events and conditions.
2. Using CAID principles for specifying and implementing intelligent agents driven by the four main causes; we call this architecture *Causal Agent*. The capability of instantiating agents and transmitting intentions is the distinctive characteristic of *CAID agents* that are on charge of generating and deploying plans.
3. Use of Electronic Institutions [4] for structuring and formalizing organizational processes. Institutional goals are managed by a CAID agent (*institutional agent*) who enables and monitors their achievement, e.g. instantiating missing agents for a given process.

4. Use of Bayesian causal models [5] for modeling agents' performance. The formal representation of controllable events is used for probabilistically modeling plan and action effectiveness. The notion of causal effect is used for quantifying indirect effects on institutional goals. Individual plans' efficiency must be accumulated in an institutional model controlled by the institutional agent.

Our methodology and the resulting framework will be applied to an information auditing scenario in an institutional corporate memory [2] as case study. Knowledge is provided by expert auditors in the form of inconsistency and correction rules which are applied to information that is fed periodically by users in an information repository. Expert auditors and users responsible for the information are requested to revise the application of those rules. An institutional agent will be on charge of two global goals: minimizing unnecessary human intervention and maximizing the confidence on the evaluation process. Individual experience must be used for modeling the efficiency of expert knowledge (auditing rules) and for adapting to environmental changes (users response) in order to contribute to global goals.

5 State of the research

I developed an ontological framework based on Aristotelian Metaphysics for representing agents and objects. Notions of essential (necessary) and potential properties are used for setting default values and knowing the capabilities of an agent.

A goal driven BDI architecture that uses an agent definition for loading sensors, actuators and plans on a generic agent architecture has been developed. Instantiation of agents is made providing: an ontology, an agent class and an efficient cause identifier. Agent goals are made explicit through the notion of *essential goals*, associated to the agent class. The final cause of an agent instantiation, i.e. its participation in an upper level plan, is transmitted as a commitment. If such commitment is contained in one essential goal the commitment is accepted. Query containment is used for this purpose.

I have incorporated a semantic layer onto a Bayesian Causal Network [3] in a formalism called SBCM. This formalism has been used for representing plans, guiding their execution and accumulating experience. Experiments on the SBCM have demonstrated that it is feasible to inhibit actions that produce undesired indirect effects.

The information auditing process of our case study has been modeled using the Electronic Institution formalism. On this framework we have implemented a mechanism for requesting the participation of agents in a scene according to a description given in terms of actual or potential properties. If there is no agent with these characteristics, the institutional agent can instantiate some agents satisfying such description. Given that actions are represented as potential properties it is possible to ask for some agent in order to request him some task.

6 Conclusions

Notions of classical philosophy have been used for formalizing the agent instantiation in terms of the four main causes. The public definition of agents allows considering the participation of current or new agents on a global plan.

The representation of institutional goals as indirect effects allows to consider them on individual action or plan selection. We still need to determine if the diffusion of these goals is sufficient for optimizing the upper level plan, and hence how it should be done.

The knowledge transference from expert auditors to autonomous agents is mediated by probabilistic learning. We expect that modeling the causal effect of individual actions over global goals will reduce the necessity of staff members for revising trustworthy rules.

Finally, self-configuration capabilities have been implemented in our framework in two different approaches: a centralized decomposition of tasks that provides a list of required agents and commitments, and a decentralized mechanism for requesting agent participation on demand.

References

1. Tomas Alvira, Luis Clavell, and Tomas Melendo. *Metafisica.* EUNSA, 2001.
2. Francisco J. Cantu, Hector G. Ceballos, Silvia P. Mora, and Miguel A. Escoffie. A knowledge-based information system for managing research programs and value creation in a university environment. In *Americas Conference on Information Systems - AMCIS*, pages 781–791. AMCIS, August 2005.
3. Hector Ceballos and Francisco Cantu. Modelling intelligent agents through causality theory. In *Proceedings of the Poster Session, MICAI 2007.* IEEE CS, 2007.
4. Marc Esteva, Juan A. Rodriguez-Aguilar, Josep Ll. Arcos, Carles Sierra, and Pere Garcia. Formalising agent mediated electronic institutions. In *Agent Mediated Electronic Commerce, LNAI 1991*, pages 29–38. Springer-Verlag, 2000.
5. Judea Pearl. *Causality. Models, Reasoning, and Inference.* Cambridge University Press, 2000.

Curriculum Vitae – Héctor G. Ceballos

ceballos@itesm.mx
http://hgceballos.wordpress.com/

I was born August 20th, 1977 at Coatzacoalcos, Ver., México.

1995 – 1999 B.S. in Computer Systems
Instituto Tecnológico de Veracruz at Veracruz, Ver.
- Ranked 2nd in college class

2001 – 2003 M.S. in Intelligent Systems
Tecnologico de Monterrey – Campus Monterrey at Monterrey, NL.
Dissertation thesis: "Ontologies handling on Multiagent Systems through an Ontology Agent applied on JITIK"
Advisor: Ramon Brena (ramon.brena@itesm.mx)
- Excellence Mention: Ranked 1st in masters class

2006 – Now PhD on Information Technologies and Communications
Tecnológico de Monterrey – Campus Monterrey at Monterrey, NL.
Dissertation thesis: "Applications of classical causality theory on the development of agents and multiagent systems"
Advisor: Francisco J. Cantú (fcantu@itesm.mx)
- Conacyt and Monterrey Tech Scolarship

I've participated on the development of several software projects at Tecnologico de Monterrey, among whose is the Institutional Repository on Research and Graduate Programs.

Publications

- Héctor G. Ceballos, Pablo Noriega and Francisco J. Cantú. Autonomic Information Auditing with Electronic Institutions. Technical Report IIIA-TR-2009-07. IIIA-CSIC. Spain. July 2009.
- Héctor G. Ceballos and Ramón Brena. Modeling Domains and Multiagents Systems with Alloy and CCalc: a Case Study on Publications Auditing. CCIR Technical Report. Tecnologico de Monterrey. México. December 2008.
- Antonio Sánchez A., Charles Hannon, Juan Pablo García-Vázquez, César García, Héctor G.Ceballos, Omar Cetina. Service Robots on a Smart Home Lab. MICAI 2008. Workshop on Service Robots. México. October 2008.
- [POSTER] Héctor G. Ceballos , Francisco J. Cantú. Modeling Intelligent Agents through Causality Theory. *6th MEXICAN INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, MICAI 2007*. MEXICO. November 2007.
- [POSTER] Héctor G. Ceballos, Francisco J. Cantú. Integrating Semantic Annotations in Bayesian Causal Models. *20*th *INTERNATIONAL WORKSHOP ON DESCRIPTION LOGICS (DL'07)*. ITALY. pp: 527-528. June 2007.

- Francisco J. Cantú, Silvia P. Mora, José A. Díaz, Héctor G. Ceballos, Sergio O. Martínez, Rosendo D. Jiménez. A knowledge-based entrepreneurial approach for business intelligence in strategic technologies: Bio-MEMS. *11*o. *Americas Conference on Information Systems*. Internacional. USA. pp: 1327-1340. August 2005.
- Francisco J. Cantú, Héctor G. Ceballos, Silvia P. Mora, Miguel A. Escoffié. A Knowledge-based information system for managing research programs and value creation in a university environment. *Americas Conference on Information Systems – AMCIS*. Internacional. USA. pp: 781-791. August 2005.
- Ramón F. Brena, Héctor G. Ceballos. Combining Local and Global Access to Ontologies in a Multiagent System. JOURNAL OF ADVANCED COMPUTATIONAL INTELLIGENCE AND INTELLIGENT INFORMATICS . JAPAN. 2005.
- Héctor G. Ceballos, Ramón F. Brena. Combinando acceso local y global a Ontologías en Sistemas Multiagentes. REVISTA IBEROAMERICANA DE SISTEMAS, CIBERNÉTICA E INFORMÁTICA . USA. 2005.
- Héctor G. Ceballos, Ramón F. Brena. Finding Compromises Between Local and Global Ontology Querying in Multiagent Systems. Ontologies, DataBases, and Applications of Semantics for Large Scale Information Systems 2004 (ODBASE04). *LECTURE NOTES IN COMPUTER SCIENCE (LNCS)*. Cyprus. pp: 999-1011. 2004.
- Ramón F. Brena, Héctor G. Ceballos. A Hybrid Local-Global Approach for Handling Ontologies in a Multiagent System. *IEEE INTERNATIONAL CONFERENCE INTELLIGENT SYSTEMS*. BULGARY. pp: 261-266. June 2004.
- Ramón F. Brena, Héctor G. Ceballos. Combining Global and Local Ontology handling in a Multiagent System. *SEVENTEENTH INTERNATIONAL FLORIDA ARTIFICIAL INTELLIGENCE RESEARCH SYMPOSIUM CONFERENCE*. USA. pp: 418-424. May 2004.
- Héctor G. Ceballos, Ramón F. Brena. Manejo híbrido local y global de Ontologías en un Sistema Multiagente. *3a. CONFERENCIA IBEROAMERICANA EN SISTEMAS, CIBERNÉTICA E INFORMÁTICA CISCI 2004*. Internacional. ESTADOS UNIDOS DE AMERICA. pp: 412-417. Julio. 2004.