

Groupe D2A2 : Droits et Devoirs d'Agents Autonomes (issu bulletin AFIA n°69)

Responsables : Olivier Boissier (EMSE), Catherine Tessier (ONERA)
Site Web : <http://d2a2.emse.fr>

L'une des particularités des SMA est de placer l'autonomie [1] [2] au cœur du comportement des agents et du système lui-même. Les agents ayant différentes tâches à réaliser, selon des contraintes diverses - par exemple suivre un protocole de négociation, suivre un schéma de coopération pour un partage de tâches avec d'autres agents, avec l'utilisateur ou un opérateur - il est important de mettre en place des mécanismes de régulation et de contrôle assurant la coopération et la cohérence du fonctionnement global des agents.

Les évolutions actuelles dans de multiples domaines (commerce électronique, jeux interactifs, robotique sociale ou collective, intelligence ambiante, etc.) renforcent cette exigence de contrôle et de régulation tout en assurant flexibilité et efficacité pour répondre à l'augmentation de l'hétérogénéité des agents, de l'ouverture et de la dynamique des applications. À terme, tout un chacun sera (est) en interaction avec des applications, des services informatisés de plus en plus enfouis dans notre quotidien et de plus en plus autonomes. Que faire en effet, lorsqu'un agent génère seul l'intention d'acheter ou de vendre un bien, lorsqu'il transmet ou traite des données personnelles de son utilisateur, lorsqu'il agit dans des systèmes tiers en y consommant des ressources et éventuellement y causant/subissant des dommages ?

Les travaux dans le domaine ont ainsi proposé différents modèles, langages et architectures pour définir ce que l'on désignera de manière générale par la suite, des *droits* et des *devoirs* d'agents autonomes. Ceux-ci se déclinent en organisations, en politiques de conversation, en normes, etc. Il est à noter qu'au-delà de ces productions scientifiques et technologiques, de nouveaux champs de réflexion tant au niveau des usages que du droit ont été également ouverts.

Les travaux et réflexions menées au sein de D2A2 s'organisent ainsi autour des trois axes suivants :

- la régulation et le contrôle des interactions entre des agents autonomes impliqués dans diverses coopérations dans un système en constante évolution afin d'assurer cohérence, efficacité et flexibilité ;
- la régulation et le contrôle du partage d'autorité entre utilisateurs/opérateurs et agents autonomes ;
- l'analyse des implications légales en termes d'anonymat, de protection de la vie privée, d'identité, de modèle de propriété de logiciel, de responsabilités afin d'ancrer ces nouveaux systèmes de transactions électroniques dans le monde réel.

Comme nous l'avons mentionné, ces travaux se déclinent en multiples propositions : (i) de langages et d'architectures, (ii) de formalismes de représentation et de raisonnement, (iii) de réflexion sur l'intelligibilité des SMA par les opérateurs ou les utilisateurs, sur les complémentarités et interactions entre les modèles du niveau multi-agents et les modèles du ni-

veau agent. Nous allons les parcourir rapidement dans ce qui suit selon les deux premières directions énoncées ci-dessus¹.

1.1.1 Régulation et contrôle

La régulation et le contrôle de systèmes d'agents autonomes ne sont pas une problématique nouvelle dans le domaine². Dès les débuts de l'Intelligence Artificielle Distribuée, il a été question d'exprimer des stratégies globales afin d'aider et de faciliter la coopération entre des résolutions entreprises localement au sein des agents (e.g. [3] [4] [5]). La prise en compte de systèmes ouverts et fortement hétérogènes a conduit à l'abandon des hypothèses de bienveillance, d'altruisme des agents. Les schémas de coopération qui étaient proposés comme une aide, ont ainsi été transformés en schémas de coopération visant à contrôler et à contraindre le comportement des agents. Ces schémas de coopération ont ainsi pris la forme de langages ou modèles d'organisations normatives (e.g. [6] [7]), de normes (e.g. [8] [9]), de lois sociales [10]. Certaines de ces propositions sont accompagnées d'outils de vérification, de validation et de cohérence. Certaines sont également complétées d'intergiciels dédiés à la gestion et au renforcement des droits et des devoirs exprimés (e.g. [11] [12]).

Se focalisant sur le niveau agent, des travaux ont proposé des modèles et des architectures d'agents capables de raisonner sur leurs droits et sur leurs devoirs dans le système (e.g. [13]). Différents types d'agents normatifs ont ainsi été proposés (e.g. [14]) afin de participer à des organisations, de raisonner sur les normes et lois en vigueur dans le système, etc.

Alors que certaines approches promeuvent la mise en place d'une partie tierce (intergiciel), en dehors des agents, pour réguler et contrôler leur comportement, certains travaux défendent une approche se référant à la notion d'ordre social [15] où les agents eux-mêmes mettent en place les modèles et les mesures pour contrôler et exclure du système les agents qui ne respectent pas les lois globales. Ce sont notamment toutes les approches relatives aux notions de confiance, de réputation (e.g. [16] [17]), etc.

Enfin, de nombreux travaux viennent compléter ces approches en s'intéressant à la dynamique et à l'évolution de ces schémas de coopération en termes d'émergence et d'évolution de normes ou d'organisations (e.g. [18] [19] [20]).

1.1.2 Partage d'autorité

Nous sommes de plus en plus amenés à déléguer ou envisager de déléguer des tâches fastidieuses, coûteuses, dangereuses, à des agents informatiques ou physiques. Cette délégation est indubitablement liée à l'autonomie conférée aux agents d'une part, et au contrôle conservé sur les agents eux-mêmes d'autre part. La définition et la mise en œuvre de ce partage d'autorité appartiennent aux défis majeurs de notre cohabitation avec des entités que nous concevons, peut-être paradoxalement, pour être à la fois autonomes et à notre service. Se reporter à [21] pour un exposé de différentes positions et illustrations de cette problématique.

Le partage d'autorité entre agents et opérateurs ou utilisateurs renvoie à la manière dont les fonctions de la boucle de décision au sein du système sont réparties ou partagées entre les

¹ Le lecteur intéressé par la dernière dimension peut se reporter à <http://www.lea-online.net/> ou <http://www.iids.org/alias> par exemple

² Ces problématiques sont abordées dans le domaine multi-agent notamment au travers de la série de workshops COIN (Coordination, Organisation, Institutions and Norms) <http://www.pcs.usp.br/~coin/>,

humains, agents logiciels ou physiques inscrits dans le système. Il se décline ainsi en différents travaux relatifs à la définition de règles de détention de l'autorité ou du pouvoir de décision [22], au contrôle de l'interaction opérateur / agent, à la définition de mécanismes de reprise en main ou de veto émanant de l'opérateur ou des agents, à l'expression de normes pour les agents et de normes pour les utilisateurs ou opérateurs [23];

La plupart du temps, un partage d'autorité implique que les agents logiciels ou physiques soient équipés d'un certain niveau d'autonomie vis-à-vis des décisions prises par le ou les humains. C'est le cas par exemple en contrôle collaboratif [24] ou dans le cadre d'initiatives mixtes [25]. Cette autonomie induit une dynamique du partage d'autorité [26] se traduisant par le fait qu'un agent (humain ou logiciel) prend le contrôle sur une action alors qu'il ne l'avait pas, de manière prévue ou non. Ceci peut être motivé par le constat que l'agent ne « sait pas faire » de manière autonome (défaillance, obstacle, danger imminent...) ou pour toute autre raison qui est propre à l'agent qui prend le contrôle et qui n'est pas forcément exprimée. D'autre part, un agent peut reprendre le contrôle de lui-même parce que ses connaissances courantes (suivi de situation) sont en contradiction avec l'action qu'un autre agent (par exemple l'opérateur) veut lui faire faire (violation de normes, ordres successifs contradictoires...) ou bien que les communications avec l'agent qui a l'autorité sont rompues.

2 Références

- [1] M. Goodrich, D. Olsen, J. Crandall, T. Palmer.: Experiments in adjustable autonomy. In: IJCAI'01 Workshop on Autonomy, Delegation and Control: interacting with autonomous agents, Seattle, WA (2001).
- [2] H. Hexmoor, C. Castelfranchi, and R. Falcone. Agent Autonomy. Boston/Dordrecht/London Kluwer Academic Publishers, 2003.
- [3] O. Boissier and J. F. Hübner and J. S. Sichman, Organization Oriented Programming: From Closed to Open Organizations, ESAW, Lecture Notes in Computer Science, Vol. 4457, pp. 86-105, Springer, 2006.
- [4] V. R. Lesser and D. D. Corkill, The Distributed Vehicle Monitoring Testbed: A Tool for Investigating Distributed Problem Solving Networks, The AI Magazine, 4(4), p. 15, Fall 1983.
- [5] V. R. Lesser, K. Decker, T. Wagner, N. Carver, A. Garvey, B. Horling, D. E. Neiman, R. M. Podorozhny, M. V. Nagendra Prasad, A. Raja, R. Vincent, P. Xuan and X. Zhang, Evolution of the GPGP/TÆMS Domain-Independent Coordination Framework, Autonomous Agents and Multi-Agent Systems, Vol. 9, Number 1-2, pp. 87-143, 2004.
- [6] J. Ferber, O. Gutknecht and F. Michel From Agents to Organizations: An Organizational View of Multi-agent Systems, AOSE, Lecture Notes in Computer Science, Vol. 2935, pp. 214-230, Springer, 2003.
- [7] J. F. Hübner J. S. Sichman and O. Boissier, A Model for the Structural, Functional, and Deontic Specification of Organizations in Multiagent Systems SBIA, LNCS, Vol. 2507, pp. 118-128, 2002.

- [8] R.Conte, C. Castelfranchi (1995). Cognitive and social action. Londra: London University College of London Press.
- [9]G. Boella, W. N. Leendert, V. der Torre, and H.Verhagen (eds.). Special Issue on Normative MultiAgent Systems, volume 12 of Computational & Mathematical Organization Theory, 2006.
- [10] Y. Shoham and M. Tennenholtz, On Social Laws for Artificial Agent Societies: Off-Line Design, Artificial Intelligence, Vol. 73, Number 1-2, pp. 231-252, 1995.
- [11] M. Esteva, J. A. Rodríguez-Aguilar, B. Rosell, and J. L. Arcos. AMELI: An agent-based middleware for electronic institutions. In AAMAS'04, pages 236-243, New York, , July 2004.
- [12] J. F. Hübner and J. S. Sichman and O. Boissier, Developing organised multiagent systems using the MOISE, IJAOSE, Vol. 1, Number 3/4, pp. 370-395, 2007.
- [13] C. Carabelea, Reasoning about autonomy in open multi-agent systems - an approach based on the social power theory" Thèse de l'ENSM Saint-Etienne, December 2007.
- [14] C. Castelfranchi and F. Dignum and C. M. Jonker and J. Treur, Deliberative Normative Agents: Principles and Architecture, ATAL'99: 6th International Workshop on Intelligent Agents VI, Agent Theories, Architectures, and Languages (ATAL), 2000, pp 364—378, Springer-Verlag
- [15] C. Castelfranchi. Engineering Social Order. In Proceedings of the First International Workshop on Engineering Societies in the Agent World (ESAW04) (p. 1 - 18). Springer-Verlag London, UK, 2000.
- [16] R.Falcone, C. Castelfranchi. Principles of trust for MAS: cognitive anatomy, social importance, and quantification, Proceedings of the International Conference on Multi-Agent Systems (ICMAS'98). In (pp. 72-79). Paris, France, 1998.
- [17] J. Sabater and C. Sierra, Review on computational trust and reputation models, Artificial Intelligence Review ,24 (1) :33-60, 2001..
- [18] G. Clair, M.-P. Gleizes, E. Kaddoum and G. Picard. Self-Regulation in Self-Organising Multi-Agent Systems for Adaptive and Intelligent Manufacturing Control. In Second IEEE International Conference on Self-Adaption and Self-Organization (SASO 2008), Venice, Italy, October 20-24, 2008.
- [19] G. Picard, C. Bernon and M.-P. Gleizes. ETTO : Emergent Timetabling Organization. In proc. of CEEMAS'05, volume 3690 of LNAI, Springer-Verlag, pages 440-449, 2005.
- [20] Verhagen H. Simulation of the Learning of Norms. Social Science Computer Review 19 (3), pp. 296–306, 2001.
- [21] A. Bouzouane, Y. Demazeau, A. Drogoul, P. Hélie, P. Taillibert, and C. Tessier, Des agents et des hommes, quels défis pour de nouveaux usages ? JFSMA 04, 2004.
- [22] B.T. Clough, Metrics, Schmetrics! How the Heck do you Determine a UAV autonomy anyway?, PERMIS Workshop, 2002

- [23] K. Myers and D. Morley, Human directability of agents, 1st International Conference on Knowledge Capture (K-CAP01), 2001
- [24] T. Fong, C. Thorpe, and C.Baur, Multi-robot remote driving with collaborative control, IEEE Transactions on Industrial Electronics 50, 2003
- [25] M. Goodrich, T. McLain, J. Crandall, J. Anderson, J. Sun, Managing autonomy in robot teams: observation from four experiments, International Conference on Human-Robot Interaction, 2007
- [26] S. Mercier, Fr. Dehais, Ch. Lesire, C. Tessier, Resources as basic concepts for authority sharing, Humans Operating Unmanned Systems (HUMOUS'08), 2008.