

Intelligence Artificielle

Russell & Norvig

3ème édition, 2010, éd. Pearson (1199 pages).

Stuart Russell travaille à l'université de Berkeley (Californie), où il est professeur au département d'informatique et directeur du centre pour l'étude des systèmes intelligents. Ancien membre du bureau exécutif de l'AAAI (*American Association for Artificial Intelligence*), il a reçu de nombreux prix scientifiques. Il est l'auteur de plus de cent articles et de plusieurs best-sellers sur l'intelligence artificielle.

Peter Norvig est directeur scientifique chez Google Inc. Auparavant, il a travaillé pour la NASA sur l'intelligence artificielle et la robotique, ainsi que pour Junglee sur l'extraction d'informations par Internet. Ancien professeur à l'université de Californie du Sud, membre de l'AAAI et de l'ACM (*Association for Computing Machinery*), il est l'auteur de nombreuses publications scientifiques et de plusieurs livres.

Supervision de la traduction: Fabrice Popineau, École Supérieure d'Électricité (Supélec).

Relecture scientifique: Laurent Miclet, École Nationale Supérieure des Sciences Appliquées et de Technologie (ENSSAT), Arthur Reutenauer, École Normale Supérieure.

4ème de couverture

Écrit par les experts de renommée mondiale, ce livre est la référence incontournable en matière d'intelligence artificielle (IA) dont il présente et analyse tous les concepts: logique, probabilités, mathématiques discrètes et du continu, perception, raisonnement, apprentissage, prise de décision et action. Sa spécificité est de présenter l'IA à travers le concept des agents intelligents. Les auteurs exposent comment un système réussit à percevoir son environnement de manière à analyser ce qu'il s'y passe, et comment il transforme la perception qu'il a de son environnement en actions concrètes.

Parmi les sujets couverts:

- * les contributions historiques des mathématiques, de la théorie des jeux, de l'économie, de la théorie des probabilités, de la psychologie, de la linguistique et des neurosciences;

- * les méthodes qui permettent de prendre des décisions lors de l'établissement d'un projet, en tenant compte des étapes à venir ;

- *les différentes manières de représenter formellement les connaissances relatives au monde qui nous entoure ainsi que le raisonnement logique fondé sur ces connaissances ;

- *les méthodes de raisonnement qui permettent d'établir des plans et donc de proposer des actions à entreprendre;

- *la prise de décisions en environnement incertain: réseaux bayésiens et algorithmes tels que l'élimination de variables et MCMC (*Markov Chain Monte-Carlo*) ;

- *les méthodes employées pour générer les connaissances exigées par les composants de prise de décision: les algorithmes de *boosting*; l'algorithme EM (*expectation-minimization*), l'apprentissage à base d'exemples et les méthodes à noyaux (machines à vecteurs support) ;

- *les implications philosophiques et éthiques de l'IA.

Chaque chapitre est illustré par de nombreux exemples et s'achève par des activités, qui vont des exercices de réflexion à des exercices de programmation, en passant par l'approfondissement des méthodes décrites, soit plus de 500 activités au total.

Cette 3ème édition tient compte des derniers développements de la matière, concernant notamment les représentations qu'un agent peut utiliser (atomique, factorisée, structurée), les environnements partiellement observables et non déterministes, les planifications contingente et hiérarchique, les modèles probabilistes du premier ordre, l'apprentissage automatique, la recherche et l'extraction d'information sur le web et l'apprentissage à partir de très grandes bases de données.

Introduction

[...]

1.2 Fondements de l'intelligence artificielle [...]

[...]

1.2.1 Philosophie

- Peut-on utiliser des règles formelles pour tirer des conclusions valides?
- Comment l'esprit émerge-t-il à partir du cerveau physique?
- D'où la connaissance provient-elle?
- Comment la connaissance conduit-elle à l'action?

Aristote (384 - 322 av. J.-C.), dont le buste apparaît sur la couverture de cet ouvrage, a été le premier à formuler un ensemble précis de lois régissant la partie rationnelle de l'esprit. Il a développé un système informel de syllogismes produisant des raisonnements valides. En principe, ce système autorisait quiconque à tirer mécaniquement des conclusions étant donné des prémisses initiales. Bien plus tard, Raymond Lulle (mort en 1315) eut l'idée qu'un artefact mécanique était capable de raisonner utilement. Thomas Hobbes (1588-1679) soutint que le raisonnement était analogue aux calculs sur des nombres que « nous ajoutons et soustrayons dans nos pensées silencieuses ». De son côté, l'automatisation du calcul était déjà bien avancée. Vers 1500, Léonard de Vinci (1452 -1519) conçut sans la construire une machine à calculer dont de récentes reconstitutions ont montré qu'elle aurait pu fonctionner. La première machine à calculer connue a été fabriquée vers 1623 par le scientifique allemand Wilhelm Schickard (1592-1635), mais la Pascaline construite en 1642 par Blaise Pascal (1623-1662) est plus célèbre. Pascal a écrit que « la machine arithmétique produit des effets qui approchent plus de la pensée que tout ce que font les animaux ». Gottfried Wilhelm Leibniz (1646-1716) a construit une machine destinée à prendre en charge des opérations sur les concepts plutôt que sur les nombres, mais ses capacités étaient assez limitées. Leibniz a pourtant surpassé Pascal en construisant une machine capable d'additionner, de soustraire, de multiplier et d'extraire des racines carrées, alors que la Pascaline ne pouvait qu'additionner et soustraire. On se mit à spéculer sur le fait que les machines pouvaient non seulement calculer, mais également penser et agir par elles-mêmes. Dans son ouvrage *Leviathan* de 1651, Thomas Hobbes suggère l'idée d'un « animal artificiel », en argumentant ainsi: « Qu'est-ce que le cœur, sinon un ressort; et les nerfs sinon autant de cordes; et les articulations sinon autant de poulies. »

C'est une chose d'avancer que l'esprit fonctionne, au moins en partie, selon des règles logiques, et de construire des dispositifs physiques qui émulent certaines de ces règles; c'en est une autre que de défendre l'idée que l'esprit lui-même *est* un tel système physique. René Descartes (1596 -1650) a été le premier à exposer clairement la distinction entre l'esprit et la matière ainsi que les problèmes qui lui sont associés. Le problème d'une conception purement physique de l'esprit tient au fait qu'elle semble laisser peu de place au libre arbitre: si l'esprit est entièrement régi par des lois physiques, alors celui-ci a autant de liberté qu'une pierre qui « décide » de tomber en direction du centre de la Terre. Descartes était un ardent défenseur de la puissance de la raison comme outil pour comprendre le monde, une philosophie que l'on nomme aujourd'hui **rationalisme**, et qui compte Aristote et Leibniz parmi ses membres.

Mais Descartes était aussi un partisan du **dualisme**. Selon lui, une partie de l'esprit humain (ou âme) était hors de la nature et soustraite aux lois physiques. En revanche, les animaux ne possédaient pas cette qualité duale et pouvaient être considérés comme des machines. Le **matérialisme** s'oppose au dualisme, selon lequel les opérations du cerveau se conforment aux lois de la physique et *constituent* l'esprit. Le libre arbitre n'est alors plus que l'aspect sous lequel l'entité qui prend la décision perçoit les choix.

La nature physique de l'esprit qui manipule des connaissances étant établie, le problème suivant consiste à définir la source de la connaissance. Le mouvement **empiriste**, qui a commencé avec le *Novum Organum* [2] de Francis Bacon (1561-1626), est caractérisé par une formule de John Locke (1632 -1704): «Il n'y a rien dans l'entendement qui n'ait d'abord été dans les sens ». Dans son *Traité de la nature humaine* (Hume, 1739), David Hume (1711- 1776) proposait ce qu'on appelle désormais le principe **d'induction**, selon lequel les règles générales sont élaborées à partir de la découverte d'associations répétées entre leurs éléments. S'appuyant sur les travaux de Ludwig Wittgenstein (1889 -1951) et de Bertrand Russell (1872-1970), le célèbre cercle de Vienne animé par Rudolf Carnap (1891-1970) a développé la doctrine du **positivisme logique**. Dans cette doctrine, toute la connaissance peut être caractérisée par des théories logiques provenant, *in fine*, de **faits d'observation** qui correspondent à des perceptions sensorielles; le positivisme logique combine donc le rationalisme et l'empirisme [3] La **théorie de la confirmation** de Carnap et de Carl Hempel (1905 -1997) a essayé d'analyser l'acquisition de la connaissance à partir de l'expérience. Dans *La Structure logique du monde* (1928), Carnap définit une procédure explicite de calcul permettant d'extraire des connaissances à partir d'expériences élémentaires. Il s'agit certainement de la première théorie de l'esprit décrivant celui-ci comme un processus calculatoire.

[2]. Le *Novum Organum* est une nouvelle version de *l'Organon* (ou instrument de la pensée) d'Aristote, lequel peut donc être considéré à la fois comme un empiriste et un rationaliste.

[3]. Dans ce cadre, tous les énoncés sont susceptibles d'être vérifiés ou invalidés soit par l'expérimentation, soit par l'analyse de la signification des mots. Comme ces règles rejetaient l'essentiel de la métaphysique - et telle était bien l'intention -le positivisme logique fut décrié dans certains cercles.

Le dernier élément de la conception philosophique de l'esprit est le lien entre la connaissance et l'action. Cette question est vitale pour l'IA, car l'intelligence requiert autant d'action, que de raisonnement. En outre, ce n'est qu'en comprenant comment les actions sont justifiées que l'on peut découvrir comment construire un agent dont les actions sont justifiables (ou rationnelles). Dans *De Motu Animalium*, Aristote a défendu le fait que les actions sont justifiées par un lien logique entre des objectifs et la connaissance du résultat des actions:

« Mais comment se fait-il que la pensée soit parfois accompagnée d'une action et parfois non, parfois d'un mouvement et parfois non? Il semble qu'il se passe la même chose lorsqu'on raisonne et qu'on produit des inférences à propos d'objets qui ne changent pas. Cependant, dans ce cas, la fin est une proposition spéculative [...] tandis que dans l'autre la conclusion produite par les deux prémisses est une action [...] J'ai besoin d'une couverture; un manteau est une couverture; j'ai besoin d'un manteau. Ce dont j'ai besoin, je dois le faire; j'ai besoin d'un manteau, je dois faire un manteau. Et la conclusion «Je dois faire un manteau» est une action. »

Page 8

Dans *l'Éthique à Nicomaque* (Livre III, 3, 1112b), Aristote approfondit ce sujet et suggère une méthode:

« Nous délibérons non sur les fins, mais sur les moyens. En effet, ni le médecin ne délibère pour savoir s'il doit guérir, ni l'orateur pour savoir s'il doit persuader. .. Mais, ayant posé en principe la fin, ils examinent comment, c'est-à-dire par quels moyens, elle sera réalisée. Et s'il se révèle possible de l'obtenir par plusieurs moyens, ils examinent par lequel elle le sera le plus facilement et le mieux. Si au contraire elle ne peut être accomplie que par un seul moyen, ils examinent *comment* elle sera obtenue par ce moyen, et ce *moyen lui-même*, par quel moyen on l'obtiendra, jusqu'à ce qu'ils arrivent à la première cause, [...] et ce qu'on trouve en dernier lieu dans l'ordre de l'analyse, c'est ce qu'on fait en premier lieu dans l'ordre de réalisation. [...] »

L'algorithme suggéré par Aristote a été implémenté deux mille trois cents ans plus tard par Newell et Simon dans leur programme GPS. De nos jours, on le qualifierait de système de

planification par régression (voir le chapitre 10).

L'analyse par objectifs est utile mais ne dit pas quoi faire lorsque plusieurs actions sont envisageables ou lorsque aucune ne permet d'atteindre complètement le but. Antoine Arnauld (1612 -1694) a fourni dans *La Logique de Port-Royal* une description correcte d'une formule quantitative pour décider de l'action à entreprendre dans de tels cas (voir chapitre 16). Dans *Utilitarisme* (Mill, 1863), John Stuart Mill (1806-1873) défend le critère de la décision rationnelle dans toutes les sphères de l'activité humaine. [théorie de la décision]. [...]
