

MODELISATION ET INTELLIGENCE ARTIFICIELLE

Bernard VICTORRI

I INTRODUCTION

D'un certain point de vue, l'Intelligence Artificielle prise dans son ensemble, en tant que discipline, peut être considérée comme une vaste entreprise de modélisation, l'une des plus ambitieuses qui soient : puisqu'il s'agit de modéliser sinon l'intégralité de la "pensée" humaine, du moins un certain nombre de ses fonctions cognitives. L'I.A. n'est évidemment pas la seule discipline à s'être donné cet objet d'études. La psychologie, la psychanalyse, les neurosciences, la philosophie, et bien d'autres, sont aussi centrées sur ce même sujet, dans des perspectives évidemment très différentes les unes des autres. Ce qui fait l'originalité de l'I.A., c'est d'être la seule dont la production soit, par essence même, de la modélisation, puisque par définition, elle implique la conception de systèmes artificiels (programmes informatiques, en fait) qui soient intelligents, et le seul sens que l'on puisse donner à ce terme qui ne soit ni trop creux ni trop trompeur, c'est bien qui ressemblent (par leurs performances ? par leurs mécanismes internes ?) au système cognitif humain, ou au moins à l'une de ses composantes. On ne peut pas prétendre aujourd'hui que l'I.A. ait atteint ses objectifs; cependant les chercheurs en I.A., bien qu'ils ne puissent exhiber de véritables modèles de l'intelligence humaine, ont le sentiment d'avoir accompli, au travers des diverses crises de la jeune histoire de leur discipline, un certain nombre de progrès décisifs. Ce sont de ces crises et de ces progrès que nous allons essayer ici de tirer des leçons qui, nous l'espérons, seront utiles pour toute entreprise de modélisation.

Il ne s'agit pas ici de faire une synthèse de l'I.A. : la discipline est trop jeune, et l'état actuel des recherches trop diversifié, pour qu'un tel travail, qui serait de toutes façons au-delà de mes compétences, soit possible et fructueux. Nous allons plutôt essayer de dégager quelques tendances, qui me semblent symptomatiques de l'évolution de l'I.A. en tant qu'entreprise de modélisation. Pour cela, rappelons que l'on peut grossièrement classer les activités de modélisation en deux catégories :

- d'un côté, les modèles qui cherchent à simuler le comportement d'un système; sous sa forme la plus caricaturale, c'est la problématique dite "de la boîte noire" : un système est caractérisé par des entrées et des sorties, et sans se préoccuper de ce qui se passe réellement à l'intérieur du système, on cherche à construire un modèle qui se comporte de la même manière, c'est-à-dire qui fasse correspondre les mêmes sorties aux mêmes entrées.

- de l'autre côté, les modèles qui cherchent à simuler les mécanismes du système à étudier. Le but est alors de reproduire le plus fidèlement possible les processus qui se produisent effectivement dans le système et de vérifier si l'on peut ainsi en retrouver les fonctionnalités.

Ces deux grandes catégories de modèles (entre lesquelles il n'est d'ailleurs pas facile de tracer une frontière nette, de nombreux modèles possédant des caractéristiques de chacune d'elles) se retrouvent en I.A. où elles constituent deux grandes tendances actuelles, que l'on appelle respectivement l'approche symbolique et l'approche connexionniste.

Nous allons donc d'abord regarder comment chacune de ces approches est née et la façon dont elle a évolué. Ensuite nous nous interrogerons sur le bien-fondé de cette classification et nous nous efforcerons de montrer que dans le cadre d'une problématique de modélisation, leurs véritables caractères distinctifs ne se situent pas dans les présupposés idéologiques qui ont présidé à leur développement mais plutôt dans le cadre mathématique dans lequel elles se placent, et par là-même dans le type de processus cognitifs qu'elles sont à même de modéliser.

II. L'APPROCHE SYMBOLIQUE

Comme nous l'avons dit, cette approche [1,2], qui a été pendant longtemps la seule à produire des résultats concrets, et qui reste prépondérante aujourd'hui, s'est donné comme but de simuler des comportements "intelligents", sans chercher a priori une quelconque ressemblance entre les techniques qu'elle utilise et les mécanismes à l'oeuvre dans le cerveau. L'objectif de l'I.A. est souvent d'ailleurs exprimé en ces termes [3]: on aura construit un système véritablement intelligent quand un individu, communiquant avec lui à distance, sera incapable de discriminer s'il a affaire à une machine ou à un humain.

Un des premiers programmes de traitement du langage "naturel", le programme ELIZA, est très symptomatique de cette vision des choses. Très célèbre à son époque (dans les années 60), il était censé simuler les réponses d'un psychiatre dans une entrevue. En fait ce programme, très simple, ne faisait que renvoyer sous forme de questions ce que venait de dire son interlocuteur (exemple : "Il fait beau généralement Pourquoi dites-vous qu'il fait beau ?) ou d'orienter la conversation quand il avait reconnu un des mots-clés qu'il avait en mémoire (Parlez-moi plus de votre mère). Evidemment ce type de programme ne peut faire illusion que l'instant de quelques répliques et n'a jamais été considéré comme "intelligent". Son principal mérite a été de mettre en évidence l'ampleur des difficultés à résoudre pour dépasser ce premier stade.

Un exemple de réel succès de l'approche "boîte noire" a été le développement de programmes d'échecs. Ces programmes [4], fondés sur des techniques simples (recherche dans l'arborescence des coups possibles d'un chemin qui maximise une certaine fonction d'évaluation, mémorisation de bibliothèques d'ouverture et d'algorithmes de fin de partie), arrivent à des performances plus qu'honorables, alors même que leur principe général de fonctionnement n'a rien avoir avec le raisonnement des joueurs humains, qui fait plus appel à des stratégies de jeu qu'à des recherches systématiques de tous les coups possibles.

Mais ce type de succès a été rare et limité à des domaines bien précis. Dans la plupart des cas, on s'est rendu compte que le seul moyen de progresser était d'analyser mieux quels étaient les éléments qui permettaient aux humains de réaliser les performances que l'on voulait imiter. Ainsi s'est imposée très vite l'idée qu'un programme d'I.A. devait disposer d'une base de connaissances, si possible incrémentable par apprentissage, dans laquelle il puisse puiser les éléments nécessaires au traitement des données. Le travail s'est donc orienté vers les techniques de représentation de connaissances, et les méthodes de déduction logique à appliquer à ces connaissances. Ainsi l'on s'est éloigné quelque peu de la philosophie de la "boîte noire". Le meilleur exemple en est la réalisation des systèmes experts [5,6], dans lesquels on cherche à modéliser la façon dont raisonne un expert humain en représentant sous formes de règles les connaissances que l'expert dit utiliser au cours de son travail pour obtenir les résultats que l'on veut simuler.

Cette évolution continue dans le même sens : de plus en plus, les chercheurs tentent de se rapprocher des processus "naturels" du raisonnement humain en prenant en compte les aspects du raisonnement qui échappent à la logique hypothético-déductive [7]: ainsi cherche-t-on à définir des "logiques floues", des logiques "non-monotones", des raisonnements sur des connaissances "par défaut", des raisonnements "à profondeur variable",... Ces différentes techniques sont autant de moyens de traquer les mécanismes à l'oeuvre dans le système cognitif humain, qui lui permettent d'être performant dans des conditions où la logique classique est inopérante : connaissances incomplètes, relatives dans le

temps et dans l'espace, présentant souvent un certain degré d'incohérence, toujours susceptibles d'être remises en question.

Bref, bien que ce ne soit pas le but explicite de sa recherche, on voit que l'approche symbolique est amenée, pour des raisons d'efficacité, à se pencher de plus en plus sur la manière dont fonctionne le système qu'elle veut modéliser.

III L'APPROCHE CONNEXIONNISTE

L'approche connexionniste, que l'on appelle aussi à juste titre neuro-mimétique, date en fait des débuts du développement de l'I.A. Cependant, malgré un démarrage prometteur, elle a connu une longue période de désintérêt, les chercheurs ayant acquis à l'époque, sans doute un peu vite, la conviction que cette voie était stérile. Ce n'est qu'au cours de ces dernières années que ce courant est revenu en force, sous l'impulsion de chercheurs en quête d'une alternative à l'approche symbolique.

L'idée au départ était de s'inspirer de la façon dont fonctionne le système nerveux en construisant des réseaux neuronaux, c'est-à-dire des réseaux dans lesquels chaque unité (neurone) est caractérisé par un degré d'activation et chaque lien entre unités par un poids positif ou négatif (synapse excitatrice ou inhibitrice). Si chaque unité représente une caractéristique du monde extérieur, un système d'apprentissage modifiant les poids permet alors au système d'apprendre à reconnaître des configurations significatives de l'environnement. Les premiers algorithmes d'apprentissage et les premières architectures de réseaux étaient très simples, mais ont permis quelques réalisations assez spectaculaires, dont la plus célèbre est certainement le perceptron, un système capable de faire de la reconnaissance de formes [8]. Les performances du perceptron étaient tout de même assez limitées et des résultats mathématiques [9] ont montré que ce type de système ne pouvait pas faire mieux, d'où le désintérêt des chercheurs. En fait ces limites n'étaient valables que pour les algorithmes simples utilisés à l'époque et on a pu démontrer ces dernières années [10] que des architectures de réseaux plus complexes, accompagnées d'algorithmes d'apprentissage adéquats permettaient de dépasser largement les performances du perceptron.

Les méthodes connexionnistes s'avèrent efficaces dans les domaines où les décisions que doit prendre le système dépendent d'un ensemble de facteurs (que l'on appelle des "micro-caractéristiques") à la manière d'une fonction de plusieurs variables presque partout continue : de petites modifications d'un facteur ne changent pas en général le résultat alors qu'une accumulation de modifications portant sur différents facteurs peut l'affecter considérablement. C'est le cas bien entendu en reconnaissance de formes, mais aussi dans bien d'autres types de "raisonnement naturel". Par exemple en linguistique, nous avons utilisé

de tels réseaux [11] pour traiter le problème des expressions polysémiques, dont le sens exact dépend de la valeur d'un grand nombre d'"indices" présents dans le contexte.

Il est sans doute encore trop tôt pour prédire la place exacte qu'occupera le connexionnisme en I.A. dans les prochaines années. Mais le point qui nous semble important pour notre propos, c'est que les progrès actuels ne sont pas dus à une meilleure simulation du fonctionnement neuronal. Au contraire les chercheurs ont abandonné très vite l'idée de simuler véritablement le système nerveux (les progrès de la neurophysiologie, qui ont montré toute la complexité des neurones auraient d'ailleurs sans doute suffi à les en décourager). Au contraire, il est apparu évident que ce qui faisait le succès de la démarche, c'était le choix de certains algorithmes, non pas parce qu'ils auraient une quelconque ressemblance avec la réalité neurophysiologique, mais parce qu'ils permettaient de manipuler avec facilité les "connaissances" que l'on voulait fournir au système.

Ainsi l'objectif initial a été, là aussi, en quelque sorte détourné et la nécessité d'améliorer les performances des réseaux conduit sans cesse le connexionnisme à s'éloigner du niveau neuronal dont il était parti.

IV. MODÉLISATION ET MATHÉMATISATION

L'évolution des différentes approches de l'I.A. montre donc que malgré des points de départ radicalement différents, les entreprises de modélisation obéissent à une loi commune qui tend à les unifier du point de vue de leur statut scientifique : c'est que toute modélisation est au fond une mathématisation du réel et en tant que telle, elle doit se soumettre aux impératifs qui régissent l'utilisation de toute théorie mathématique dans n'importe quel domaine scientifique. D'une part, on ne peut parler de mathématisation du réel que si l'on peut établir une certaine correspondance entre les processus réels et les objets mathématiques que l'on définit. Mais d'autre part, ces objets mathématiques sont de nature radicalement différente des entités réelles et les opérations mathématiques qui les manipulent n'ont pas, en général, de correspondant "naturel" immédiat. Le point géométrique ou le champ électrique ont une définition et des propriétés intrinsèques, et les équations de Maxwell n'admettent pas de description en dehors du langage mathématique.

Ceci ne signifie pas qu'il n'y ait pas des différences importantes entre les deux approches : mais elle sont à rechercher dans le type de mathématiques qui sont utilisées, et non pas dans la démarche épistémologique. De fait l'approche symbolique s'appuie sur la logique et la théorie des graphes, c'est-à-dire sur des mathématiques essentiellement discrètes, alors que les réseaux connexionnistes s'inscrivent dans le cadre de la géométrie différentielle et de la théorie des systèmes dynamiques, c'est-à-dire sur les mathématiques du continu. Ce dernier point demande

peut-être à être explicité : on peut montrer [12] qu'en se plaçant dans l'espace constitué par les activités des unités de sortie, le processus de stabilisation d'un réseau peut être décrit comme une dynamique créée par les valeurs des unités d'entrée sur cet espace et l'on peut, dans ce cadre, caractériser le réseau par les différentes formes de dynamiques qui lui sont associées.

Le problème qui se pose est donc de savoir si l'on a avantage à représenter les processus cognitifs dans le cadre du discret ou du continu. Mais une fois posé ainsi, la réponse semble évidente : certains processus se prêtent parfaitement à une représentation discrète et le passage par le continu est un détour inutile et inefficace, alors que d'autres réclament intrinsèquement du continu et ne sauraient en permettre l'économie.

Comment reconnaître ces deux classes de processus? J'émettrais volontier l'hypothèse que le critère décisif se situe au niveau du rapport local/global : tout processus pour lequel l'information locale ne prend sa signification qu'au niveau global nécessiterait le continu. On comprend alors que dans un même domaine puissent coexister des modèles connexionnistes et symboliques. Par exemple, en reconnaissance de forme, si l'on prend comme information locale des caractéristiques qui ont déjà une signification de forme (angles et sommets de polygones, par exemple) des techniques symboliques sont pleinement efficaces alors que si l'on prend l'information "brute", un réseau connexionniste est plus performant. Cette hypothèse demanderait bien entendu à être approfondie. Mais il me semble que c'est en poussant dans cette direction que l'on pourrait utilement discriminer les différentes méthodes de l'I.A.

V. CONCLUSION

La réflexion présentée ici ne saurait prétendre faire le tour de la problématique de la modélisation : L'adage dites-moi quelles mathématiques vous utilisez, je vous dirai quel type de modèle vous construisez est bien sûr quelque peu simpliste et unilatéral. Mais il a l'avantage de mettre l'accent sur la spécificité de la modélisation dans l'activité scientifique plus générale de construction de théories. Il est clair pour tout le monde qu'il existe des théories qui ne sont pas des mathématisations d'un domaine (la psychanalyse par exemple). Mais ce qui est peut-être moins clair, c'est que le choix de fabriquer un modèle dans le domaine que l'on étudie implique automatiquement le choix de construire une théorie mathématique de ce domaine : c'est donc un choix épistémologique important.

Ce point de vue permet de clarifier bien des malentendus sur le bien fondé de tel ou tel modèle. C'est ainsi qu'en ce qui concerne l'I.A., on peut, à mon avis, renvoyer dos à dos les protagonistes du débat "philosophique" qui s'est développé ces derniers temps outre atlantique

entre tenants de l'approche symbolique et défenseurs du connexionnisme. En effet, si l'on admet que d'un point de vue épistémologique les deux approches relèvent d'une même démarche scientifique, à savoir la production de théories mathématiques de processus cognitifs, on ne saurait valider ou invalider l'une des approches sur la base d'arguments philosophiques. Seule la pertinence des mathématiques utilisées et l'efficacité des modèles construits doit entrer en ligne de compte.

VI BIBLIOGRAPHIE

- [1] NILSON N.J. : Principles of Artificial Intelligence , Springer-Verlag, 1980.
- [2] WINSTON P.H. : L'Intelligence Artificielle , Interéditions, 1987.
- [3] FEIGENBAUM et FELDMAN : Computers and Thought, McGraw-Hill, 1963.
- [4] NEWBORN M. : Computer Chess , Academic Press, 1974.
- [5] SHORTLIFFE E. : Computer-based medical consultations : MYCIN , American Elsevier, 1976.
- [6] DAVIS et LENAT : Knowledge-based systems in Artificial Intelligence, McGraw-Hill, 1982.
- [7] PRC-GRECO : Actes des journées nationales de Toulouse, Teknea, 1988.
- [8] ROSENBLATT F. : Principles of neurodynamics, Spartan, 1962.
- [9] MINSKY et PAPERT : Perceptrons, MIT Press, 1969.
- [10] RUMELHART et McLELLAND : Parallel Distributed Processing, MIT Press, 1986.
- [11] FUCHS et VICTORRI : Une approche topologique de l'ambiguïté, Actes du colloque de l'ARC de Toulouse, 1988.
- [12] VICTORRI B. : Polysémie et théorie des catastrophes, en préparation.