Auto-apprentissage et indiscernabilité.

La robotique autonome et la vie artificielle

Michel Troublé – directeur de recherche en robotique et intelligence artificielle.

Pour que soit assurée la durabilité d'un robot *autonome*, artificiellement *vivant*, son analyse fonctionnelle montre que pour chaque objet du monde, aux propriétés infiniment variées, que celui-ci perçoit à l'aide de différents capteurs, il faut que son contrôleur (son « cerveau ») sache choisir adéquatement les actions à accomplir par le truchement de ses actionneurs (système de locomotion, pinces de préhension,...).

En d'autres termes, le contrôleur d'un robot *autonome* doit être capable d'*auto-apprendre*, c'est-à-dire de créer des *catégories cohérentes* des objets perçus par ses capteurs. Par exemple, pour un robot d'exploration terrestre sensible à toute élévation de température qui peut le détruire, il doit impérativement : *fuir* de la lave en fusion, *fuir* un feu de forêt, *fuir* une nappe de pétrole enflammée,... Soit la *catégorie cohérente d'action* ou *concept* {*fuir* tous les objets chauds}, constituée par l'ensemble des objets perçus par le robot devant produire la même action de *fuir* qui pérennise la structure du robot.

Le *concept* est un processus opératif qui est donc essentiel à l'existence même d'un robot autonome, artificiellement vivant.

Du point de vue du mécanisme de *décision* dont le contrôleur du robot doit être muni afin d'effectuer des actions adéquates qui assureront sa pérennité, cette capacité de créer des *catégories cohérentes d'actions* implique logiquement que les différents objets perçus par le système soient physiquement *discernables* les uns des autres. Dans le cas contraire, ces *décisions* ne pourraient être en effet que prises au hasard ce qui serait assurément antinomique de la capacité attendue de former des *catégories cohérentes* qui fonderaient son *autonomie*, son état de *vie*.

Que les objets du monde perçus par les capteurs du robot soient physiquement *discernables* afin que celui-ci puisse, par exemple, *fuir* systématiquement les objets chauds et *aller* au contraire vers des sources d'énergie supposées froides, semble aller naturellement de soi. Mais cela pose en réalité un problème épistémologique majeur.

En s'appuyant sur la théorie formelle de la « Reconnaissance des formes¹ » qui porte sur l'identification des formes d'objets à partir de leurs paramètres caractéristiques afin de prendre des *décisions* dépendant des catégories attribuées à ces formes, on peut en effet démontrer le trait essentiel suivant :

Les différentes formes des objets qui sont perçues par un système physique (ou physico-chimique) lors d'un processus de mesure, sont physiquement *indiscernables* par sa partie opérative ou actionneur.

Ce 'théorème d'indiscernabilité', pour faire court, s'applique à tous les niveaux de la matérialité – macroscopiques ou microscopiques/quantiques –, indépendamment des lois physiques qui régissent ces domaines. Il peut être établi² en analysant la nature des liaisons physiques qui doivent être mises en place entre le *capteur* et l'*afficheur* d'un appareil de mesure dont la fonction essentielle est de déterminer les propriétés spécifiques des objets du monde avec lesquels un système physico-chimique interagit.

Il en résulte que l'autocréation d'un *concept* comme {*fuir* tous les objets chauds} qui se caractérise par la perception par le robot d'un ensemble d'objets de formes diverses qui doivent provoquer une *action* commune qui assure la pérennité de sa structure, est physiquement impossible.

¹ Satosi Watanabe - *Pattern recognition, human and mechanical*, John Wiley & Son, 1985.

² Voir Annexe

Mais que peut-on dire alors de la technique *d'auto-apprentissage* du 'deep learning' qui est réputée comme étant capable de découvrir par elle-même le *concept* {chat} par le seul fait qu'un neurone **Nc** du système se trouve spécifiquement activé lorsqu'on montre un chat au système de perception du réseau après une longue phase d'apprentissage non-supervisée durant laquelle ont été présentées une multitude d'images comprenant toutes sortes d'objets, dont des chats.

Un *concept*, élément de base qui fonde la cognition de toutes les structures vivantes ou artificielles comme les robots autonomes, est l'ensemble des différentes formes des objets perçues qui doivent produire une même *action* dont le rôle est d'assurer la pérennité d'une structure matérielle. Ainsi, en l'absence d'une *action* spécifique, la seule activation du neurone **Nc** dans l'expérimentation effectuée avec la technique du 'deep learning' n'a donc aucune signification opérative.

Mais eu égard le 'théorème d'indiscernabilité', les états activés (dont **Nc**) ou non des neurones du système 'deep learning' sont strictement *indiscernables* du point de vue d'un actionneur. C'est un technicien qui observant le réseau décide, utilitairement, que le neurone **Nc** a un état différent des autres neurones, et *active* alors un dispositif adéquat qui affiche, par exemple, l'image d'un chat et non celle d'un cheval.

La technique d'auto-apprentissage du 'deep learning' ne permet donc pas l'auto-découverte d'un quelconque *concept* comme celui du {chat}. Le 'deep learning' n'est pas une explication pertinente à la capacité opérative du processus de cognition dont sont dotées les structures vivantes, capacité opérative qui leur permet d'assurer leur pérennité, leur état de vie.

- Annexe -

'théorème d'indiscernabilité' :

Pour effectuer la mesure d'un observable **O** sur un objet **A**, il faut que cet objet interagisse avec un dispositif technique donné **Ap** qui a la particularité de se mettre dans un état final spécifique unique **E** lorsque l'interaction est complète.

Considérons un dispositif de mesure thermométrique constitué des éléments suivants : un *capteur* (un thermomètre), un *afficheur* (ou *actionneur*) et des *liaisons physiques* entre le capteur et l'actionneur. Le monde sur lequel portent les mesures thermométriques est supposé constitué des seuls deux objets **A** et **B** dont le premier est, par exemple, de la lave en fusion, il est *chaud*, et le second un morceau de glace, il est *froid*. Les mesures sont ainsi relatives au seul observable **P**, la *température*.

Généralement, si **P** est le nombre d'observables, il y a $N = 2^P$ propriétés possibles. En l'occurrence, avec le seul observable *température*, soit $P = 1$, il en résulte qu'il existe $N = 2^P = 2^1 = 2$ propriétés différentes. À savoir, les propriétés 'chaud' et 'froid' qui caractérisent respectivement les objets **A** et **B**.

À un état donné (*chaud* ou *froid*) de l'entité **A** ou **B** qui fait l'objet de la mesure doit correspondre un *état unique* du capteur thermométrique. Ainsi, avec un thermomètre en tant

qu'élément sensible **A_p** à l'observable *température*, cet état est représenté par la longueur de la colonne de mercure qui est fonction de la température de l'objet sur lequel porte la mesure.

En considérant que le « zéro de la règle » qui permet de déterminer la longueur de la colonne de mercure se trouve au niveau du bulbe du thermomètre, deux cellules photoélectriques **CA** et **CB** – seulement sensibles à la forme spécifique du ménisque de mercure – sont positionnées en deux points précis le long du capillaire. Les localisations des deux cellules **CA** et **CB** correspondent aux deux positions possibles atteintes par le ménisque.

Le signal de sortie **SA** pour la cellule **CA**, ou le signal **SB** pour la cellule **CB**, est égal à **1** lorsque le ménisque se trouve devant la cellule, et **0** dans le cas contraire. Ne sont pris en compte que les positions finales du ménisque de mercure, les positions transitoires sont inintelligibles.

Lorsque l'objet **A** se trouve devant le bulbe du thermomètre et que la mesure est complète (la colonne de mercure est stabilisée), seule la cellule photoélectrique **CA** est finalement activée, soit un signal de sortie **SA** = 1, avec **SB** = 0. Pour l'objet **B**, c'est seulement la cellule **CB** qui est activée, soit un signal de sortie **SB** = 1, avec **SA** = 0.

Les *liaisons physiques* qui s'établissent entre le capteur et l'afficheur doivent répondre à la question suivante : comment passer d'un système *capteur* qui possède deux sorties distinctes **SA** et **SB**, à l'entrée unique d'un *afficheur*.

Logiquement, il y a 4 et seulement 4 combinaisons physiques possibles entre les deux sorties **SA** et **SB** qui conduisent à une sortie unique, soit : **0**, **SA**, **SB**, {**SA ou SB**} (combinaisons déclinées à partir du développement exhaustif d'une 'table de vérité' carrée fondée sur les deux entrées **SA** et **SB** qui peuvent prendre les valeurs 0 ou 1).

La combinaison {**SA et SB**}, au moyen de l'opérateur « *et* », est logiquement toujours égale à 0 puisque les cellules photoélectriques **CA** et **CB** ne peuvent pas être simultanément activées lors de la présentation des objets **A** ou **B** devant le thermomètre.

Généralement, il y a $M = 2^N$ combinaisons possibles établies à partir des **N** propriétés possibles (ce sont les **N** lignes qui résultent des différentes combinaisons possibles de 0 et de 1 d'un tableau possédant **N** colonnes). En l'occurrence, le fait qu'il existe **N** = 2 propriétés *chaud* et *froid* attachées respectivement aux objets **A** et **B**, entraîne qu'il y a bien $M = 2^N = 2^2 = 4$ combinaisons possibles : **0**, **SA**, **SB**, {**SA ou SB**}

En considération du contexte expérimental, les 4 liaisons possibles entre le capteur et l'afficheur sont donc les suivantes :

- une liaison **L0**, toujours inopérante car reposant sur une combinaison des sorties **SA** et **SB** qui est toujours égale à 0.
- une liaison **L1**, attachée à la sortie **SA**.
- une liaison **L2**, attachée à la sortie **SB**.
- une liaison **L3**, attachée à la sortie composite {**SA ou SB**}.

Les quatre liaisons **L0**, **L1**, **L2**, **L3** entre le capteur et l'afficheur étant établies, on réalise alors les tâches suivantes :

- (1) l'objet **A** est placé devant le bulbe du capteur thermométrique : seule la cellule photoélectrique **CA** est activée, d'où **SA** = 1 et par là même {**SA ou SB**} = 1. Les liaisons **L1** et **L3** sont donc simultanément opérantes.
- (2) l'objet **B** est placé devant le bulbe du capteur thermométrique : seule la cellule photoélectrique **CB** est activée, d'où **SB** = 1 et par là même {**SA ou SB**} = 1. Les liaisons **L2** et **L3** sont simultanément opérantes.

Finalement, les objets **A** et **B** sont donc respectivement vérifiés par les liaisons **L1** et **L2**, et à ce titre, ces objets sont *différents*. Dans le même temps, les objets **A** et **B** sont simultanément vérifiés par la liaison **L3**, et à ce titre, ils sont aussi *identiques*.

Généralement, parmi les $M = 2^N$ combinaisons possibles qui résultent de l'existence de N propriétés, il y a $E_1 = M/4$ combinaisons qui vérifient *spécifiquement* chacune des N propriétés, et $E_2 = M/4$ combinaisons qui vérifient *simultanément* deux quelconques de ces propriétés.

En l'occurrence, le fait qu'il existe deux propriétés *chaud* et *froid* attachées respectivement aux objets **A** et **B**, entraîne qu'il existe : (1) $E_1 = M/4 = 2^N/4 = 2^2/4 = 1$, soit une liaison **L1** attachée à la sortie **SA** et une liaison **L2** attachée à la sortie **SB** vérifiant *respectivement* les objets **A** et **B**, (2) $E_2 = M/4 = 2^N/4 = 2^2/4 = 1$, soit une liaison **L3** attachée à la sortie composite **{SA ou SB}** vérifiant *simultanément* les objets **A** et **B**.

Étant à la fois *différents* et *identiques*, les objets **A** et **B** perçus par un système de mesure sont strictement *indiscernables* du point de vue de sa partie opérative ou actionneur.