De quelle variabilité sont capables les réseaux de neurones artificiels ?

Frédéric Alexandre

De quelle variabilité sont capables
les réseaux de neurones artificiels?

Frédéric ALEXANDRE
LORIA/INRIA-Lorraine
BP 239
F-54506 Vandoeuvre
falex@loria.fr

A priori, lorsque l'on parle d'invariants et de variabilité, le chercheur dans le domaine des réseaux de neurones artificiels (RNA) est tenté de dire que ses outils préférés sont plutôt bien équipés pour manipuler ou rendre compte de ces phénomènes importants de la cognition. Puis, en creusant un peu plus ce problème, on se rend compte que la question n'est pas aussi facile qu'il n'y parait à première vue. C'est cette réflexion que je vais tenter de développer ici en présentant tout d'abord les aptitudes intrinséques des RNA face à ces phénomènes, puis leurs limitations et en proposant de nouvelles voies de recherche pour les dépasser.

1 Extraction et représentation des invariants par RNA

Les RNA sont des réseaux d'unités simples, interconnectées, effectuant des calculs numériques à partir de leurs entrées, pour évaluer leurs sorties qui seront à leur tour intégrées par d'autres unités. En plus de ces propriétés originales de calcul numérique et distribué, les RNA se distinguent également par une troisième propriété fondamentale: l'adaptativité qui les rend capables d'apprentissage. Cette rapide description permet dès maintenant d'expliquer comment des RNA peuvent représenter et extraire des invariants à partir de données numériques qui leur sont présentées en entrée. Pour cela, considérons simplement un neurone N_i, recevant en entrées les activités de n autres neurones N_j, avec $j \in \{1, \cdots, n\}$, l'influence de ces activités étant pondérée par des poids, aussi appelés coefficients synpatiques, W_{ij}.

La loi de fonctionnement du neurone N_i consiste alors à définir la fonction f qui évalue l'activité du neurone N_i en fonction des activités des neurones N_j et des poids W_{ij}.

$$e_i = f(e_j, W_{ij})_{j \in \{1, \cdots, n\}} \quad (1)$$

La forme précise de cette loi de fonctionnement peut changer d'un modèle de RNA à un autre, mais il est important de souligner ici que, dans tous les cas, cette loi permet au neurone N_i de représenter un invariant du monde extérieur que l'on appellera, selon les modèles, représentant, prototype, frontière, etc.

A son tour, la loi d'apprentissage définit les critères de modifications des poids W_{ij} permettant d'extraire, à partir d'exemples, les invariants utiles à la tâche réalisée. Illustrons maintenant la représentation et l'extraction de ces invariants pour les principaux modèles de RNA.

1.1 Réseaux directs à couches

Pour les réseaux directs à couches, de type perceptron, la loi de fonctionnement s'écrit comme la somme des entrées, pondérée par les coefficients synaptiques, pour laquelle une fonction à seuil g détermine le niveau d'activité du neurone N_i.

$$e_i = g(\sum_{j \in \{1,\cdots,n\}} W_{ij}.e_j) \tag{2}$$

Dans l'espace des entrées e_j et pour des formes simples de g, cette équation représente un hyperplan séparant cet espace en deux parties et, pour une entrée donnée e_j, l'activité du neurone N_i va donc simplement indiquer de quel coté de l'hyperplan se trouve cette entrée.

Ces réseaux à couches de tels neurones sont généralement utilisés dans des tâches de classification (discrimination) ou d'approximation de fonctions (régression) où l'apprentissage consiste à positionner les frontières portées par les neurones de manière à réaliser au mieux la classification. Pour cela, dans le cadre d'un apprentissage supervisé, le calcul de l'erreur entre l'activité réelle et l'activité désirée des neurones permet de modifier leurs poids et par là même la position des frontières séparatrices. Ainsi, ces frontières, qui sont les invariants du monde extérieur utiles pour cette tâche, sont extraits par apprentissage et représentés par les poids des neurones.

1.2 Cartes auto-organisatrices

Dans le cas des cartes auto-organisatrices (et d'une manière similaire pour les modèles de la théorie de la résonance adaptative de Grossberg), la loi de fonctionnement s'écrit :

$$e_i = \sum_{j \in \{1,\cdots,n\}} (e_j - W_{ij})^2 \tag{3}$$

Autrement dit, le neurone N_i calcule une distance entre son vecteur poids et l'entrée e_j qui lui est soumise. Le vecteur poids (et à travers lui le neurone N_i) peut ainsi être vu comme une entrée virtuelle, ou encore un prototype ou un représentant d'une telle entrée.

L'apprentissage non supervisé consiste, pour une entrée donnée, à sélectionner le (ou les) neurone le plus proche de cette entrée et à modifier ses poids de manière à le rendre encore plus proche de cette entrée. Devant une distribution de points, présentés en entrée, un ensemble de neurones va donc se répartir de manière à représenter au mieux cette distribution de points, en se plaçant au centre des groupes de points les plus significatifs. Ainsi ces neurones deviennent des représentants ou des prototypes de ces classes de points. Selon un autre point de vue, ils effectuent une opération de filtrage en définissant, par leurs poids, des filtres prototypes des configurations d'entrées les plus représentatives.

1.3 Réseaux récurrents

Dans les réseaux récurrents, comme le modèle de Hopfield, les connexions ne sont pas destinées à recueillir les entrées, qui sont fournies une par une aux neurones, mais à interconnecter totalement les neurones entre eux. La loi de fonctionnement, similaire à celle des réseaux directs, va alors apprendre des configurations du réseau (et non pas des entrées), comme des états stables. Ensuite, devant un état initial quelconque, le réseau saura osciller de manière à converger vers l'état stable appris le plus proche. Ces états stables appris sont aussi appelés prototypes et peuvent être vus comme des invariants de configuration du réseau, utilisé comme mémoire associative, ou mémoire adressable par son contenu.

1.4 Discussion

Nous avons montré comment des RNA simples, destinés à des opérations de traitement statistique de données, pouvaient en extraire par apprentissage et représenter en leur sein, les invariants permettant de réaliser ces opérations de filtrage, de classification, de mise en correspondance ou d'association.

Partant de ces modèles élémentaires, plusieurs auteurs (Kohonen, Grossberg, Rubner) ont souligné la proximité entre ces opérations statistiques et l'extraction, par le système nerveux des êtres vivants, des invariants du monde extérieur qui leur permettent de se créer une représentation interne du monde extérieur. On cite ainsi volontiers les cellules sélectives à l'orientation des stimuli visuels perçus du cortex visuel et les filtres similaires obtenus par les cartes auto-organisatrices de Kohonen.

S'il est ainsi possible de s'accorder pour souligner l'importance primordiale de cette notion d'invariants, aussi bien pour la cognition que pour le traitement statistique de données et pour admettre, en première approximation, une certaine similarité entre les mécanismes biologiques et statistiques aussi bien pour la représentation que pour l'extraction de ces invariants, il est difficile de se satisfaire aujourd'hui de ces lois mathématiques pour modéliser ces phénomènes cognitifs.

En effet, les connaissances sur ces phénomènes cognitifs et en particulier sur les mécanismes biologiques sous-jacents permettent maintenant de préciser la nature de ces invariants et de ne plus les réduire à la simple extraction de primitives visuelles. Plus précisément, l'étude du comportement de populations de neurones (Koechlin) indique que pour décider de leur activation, les neurones ne se fondent pas seulement sur l'information montante (feed-forward) qui leur parvient de l'extérieur ou des couches inférieures, mais aussi du comportement de leurs voisins obtenu par des connexions latérales ou encore d'indices assimilables à des intentions, des motivations ou d'autres informations sensorimotrices provenant d'autres aires corticales par des liens récurrents (feed-back).

Cette nouvelle vue sur la nature des invariants ainsi construits par intégration d'informations de nature différente a plusieurs conséquences. Au niveau cognitif et comportemental, elle indique que les invariants qui structurent la cognition ne sont pas simplement perceptifs, mais qu'ils se fondent également sur d'autres aspects de la représentation interne, comme les croyances, les représentations motrices, les drives, etc. On pourrait y retrouver là la notion d'affordance (Gibson). Au niveau neuronal, cela met au premier rang l'unité fonctionnelle de la population neuronale (plutôt que le neurone isolé) pour représenter les invariants, avec l'idée de synchronisation entre neurones ou encore d'intégration progressive de l'information dans le temps (d'abord feed-forward puis latérale puis récurrente) qui donne une dimension temporelle à l'invariant, mais aussi un aspect dynamique bien supérieur à la simple image du filtre. Au niveau de la modélisation, ces indices épars restent à être

intégrés de manière cohérente et opératoire dans des modèles informatiques. Les modèles formels issus des mathématiques n'ont pas su donner pour le moment des règles d'apprentissages permettant d'intégrer des informations de nature et d'échelle temporelle différentes, provenant de liens typés (feed-forward, latéraux et récurrents). Plusieurs modèles d'inspiration biologique ont été proposés et restent à être précisés.

2 La variabilité des RNA

Pour parler du comportement des RNA vis à vis de la variabilité, il convient tout d'abord de définir la notion de variabilité. Si l'on s'intéresse à la variabilité du signal d'entrée, on dira surement alors que les RNA savent bien prendre en compte cette variabilité. Il n'en sera certainement pas de même si l'on s'intéresse à la variabilité de l'expression des stratégies.

2.1 Variabilité du signal d'entrée

La prise en compte de ce type de variabilité renvoie à une des propriétés les plus connues des RNA, la robustesse, qui peut se décliner de différentes manières. Pour des modèles avec des sorties discrètes (Hopfield, Kohonen, perceptrons utilisés en discrimination), on constate généralement que des entrées proches vont donner lieu à des sorties identiques ou proches. Selon les modèles, on parlera de bassins d'attraction, de propriétés topologiques ou encore de généralisation. Dans le cas de modèles à sorties continues (perceptrons utilisés en régression), on observera aussi des sorties obtenues par interpolation à partir d'exemples proches vus au cours de l'apprentissage. Tous ces phénomènes donnent des réseaux robustes à l'introduction de bruit ou d'imprécisions dans les données qui leur sont fournies en entrée.

2.2 Variabilité des réponses

En revanche, ces mêmes réseaux sont très peu capables de rendre compte de variabilité de l'expression de réponses et plus généralement de stratégies, que l'on peut ranger sous le terme de vicariances (Gibson). Ceci peut simplement s'expliquer par le fait que les réseaux sont généralement construits pour le réalisation d'une seule fonction et non pas pour l'articulation de plusieurs, ce qui pourrait donner lieu à plus de souplesse. Lors de la réalisation d'une

seule fonction, la seule variabilité observée en sortie correspond généralement à l'utilisation d'une fonction aléatoire qui permet l'exploration d'une partie de l'espace des solutions.

L'alternative consiste à étudier des réseaux modulaires dont l'architecture même peut donner lieu à l'expression de comportements plus différenciés, comme on l'observe par exemple dans des modèles comme les mixtures d'experts (Jordan). Dans ce domaine, la principale limitation tient au type de fonctions réalisées par ces réseaux. On reste en effet au niveau de fonctions élémentaires de traitement de données (ex: classification) et on évite le plus souvent des tâches plus complètes et plus réalistes mettant en oeuvre plusieurs fonctions (ex: exploration de l'environnement nécessitant localisation, reconnaissance de lieux, planification, etc.) pour lesquelles le type de variabilité qui nous intéresse ici s'exprimerait mieux. Aussi, sans qu'il soit encore possible de préjuger des compétences des RNA sur ce point, l'avancement dans ce domaine se confond avec le développement de réseaux modulaires pour de telles tâches plus cognitives et plus écologiques. Là aussi, les RNA d'inspiration biologique ont ouvert la voie, mais il reste encore beaucoup de progrès à faire pour la mise au point de réseaux capables de produire de tels comportements dynamiques et, qui plus est, de travailler sur leur variabilité.