

Avant Propos

Pourquoi simuler des réseaux de neurones sur un ordinateur ?

Partir de la formalisation mathématique du fonctionnement d'un neurone biologique, et trouver une solution viable - souvent efficace - à de nombreux problèmes (classification, prévision, compression, contrôle, etc.), justifie tout l'intérêt porté aux réseaux de neurones artificiels. Le même état d'esprit conduit tout naturellement à s'intéresser au parallélisme, c'est-à-dire à la capacité de faire plusieurs choses en même temps. En effet, le cerveau fonctionne en parallèle à plus d'un titre, simuler cette capacité est donc à la fois naturel et prometteur.

Bien sûr, cette thèse avance d'autres bonnes raisons, mais l'inspiration biologique reste à la base des réseaux neuro-mimétiques. S'inspirer de la nature a souvent aidé les sciences. Si notre planète était dépourvue d'oiseaux, aurions-nous mis des ailes à nos machines volantes ? Mais une simple montgolfière rappelle que la nature ne détient pas toutes les solutions. Une des plus grandes inventions de l'humanité est inexistante dans la nature : que serait notre société sans la roue ? Il faut donc se garder de voir dans le connexionnisme une solution miracle et universelle. A contrario, un argument souvent opposé au connexionnisme est qu'une solution algorithmique peut toujours faire au moins aussi bien. Le fait que l'on soit obligé de simuler les réseaux de neurones artificiels sur ordinateur ne permet pas d'en douter. Mais s'il est concevable de penser qu'un algorithme suffisamment sophistiqué puisse un jour « égaler le cerveau », il est plus difficile de l'écrire. Le connexionnisme est donc une source d'inspiration pour apprendre à utiliser nos ordinateurs.

Remerciements

Je tiens à remercier ici Hélène Paugam-Moisy, pour m'avoir donné l'opportunité de faire ce travail et pour avoir toujours su trouver du temps à m'accorder. Merci aussi à Michel Cosnard pour sa confiance et son « positivisme ». Merci à Frédéric Alexandre et à Christian Jutten pour avoir bien voulu rapporter ma thèse, ainsi qu'à Bernard Amy et à Djamel Zighed pour m'avoir fait l'amitié d'être dans mon jury.

Merci à mes amis pour leur soutien, notamment Alice, Anahi, Anne, Audrey, Bérengère, Catherine-Charlotte, Cécile, Geneviève, Isabelle, Nathalie, Pauline, Regina, Claude, Cyril, Denis, Francis, Jean-Marc, Julio, Laurent, Richard et Sylvain.

Enfin, merci à mes parents, à mon frère Eric et à ma belle sœur Patricia, pour leurs encouragements.

Résumé

Le connexionnisme permet de résoudre des problèmes en simulant des réseaux de neurones. La mise en oeuvre d'un réseau de neurones artificiels impose de lourds calculs et motive l'utilisation des machines les plus puissantes : les ordinateurs « MIMD à mémoire distribuée ». Cette architecture - dite à gros grain - se traduit par de puissants processeurs équipés d'une mémoire locale. Les processeurs communiquent *via* un réseau d'interconnexion plus ou moins complexe.

Cette thèse s'intéresse plus particulièrement à la parallélisation d'un « classifieur incrémental ». Ce modèle connexionniste discrimine un ensemble de données en classes, des cellules sont ajoutées en fonction des besoins. La nature dynamique du modèle rend plus difficile la parallélisation, mais les solutions apportées sont d'autant plus générales. Une première parallélisation distribue l'espace d'entrée du réseau entre les processeurs. Le classifieur parallélisé suit parfaitement le comportement séquentiel. L'accélération maximale est d'autant plus grande que la dimension de l'espace d'entrée est élevée.

Un réseau de neurones se distingue d'un algorithme numérique par une phase d'apprentissage sur des exemples. Plusieurs parallélisations, dites modulaires, sont proposées en parallélisant non plus le réseau, mais l'apprentissage. Le comportement parallèle du modèle diffère du comportement séquentiel mais le niveau de performance en classification est maintenu. Une spécialisation des modules permet d'atteindre une accélération optimale (linéaire). Une accélération supra-linéaire met en lumière l'intérêt d'un apprentissage « cloisonné » : chaque module travaille sur un sous-problème cohérent. L'efficacité de l'approche modulaire a conduit au développement d'une version asynchrone adaptée à un réseau de stations de travail. Cette nouvelle version rend tout le travail réalisé indépendant de la machine cible en utilisant une machine virtuelle.

Au-delà d'une utilisation en ingénierie, le connexionnisme permet de modéliser et de comprendre le vivant dans le cadre des sciences cognitives. Le classifieur est notam-

ment mis à contribution pour simuler un phénomène « d'amorçage de répétition » bien connu des psychologues. Par ailleurs, l'expérience accumulée sur le classifieur est mise à profit pour analyser les résultats d'expériences olfactives réalisées auprès de sujets humains. Ce travail permet de valider une hypothèse faite par les neuropsychologues sur l'architecture fonctionnelle du système perceptif olfactif. De plus, l'étude montre que le vocabulaire courant est peu adapté à la discrimination de stimuli olfactifs. La base d'exemples, difficile à apprendre, issue des réponses des sujets humains, permet d'améliorer le classifieur en évitant un « sur-apprentissage » au moyen de la stratégie de cloisonnement. La thèse s'achève sur la perspective de modéliser les processus de mémorisation humains en associant des classifieurs incrémentaux à une mémoire associative. Ce système, hautement modulaire, ouvre un champ d'application très large.

Introduction

L'étude des réseaux de neurones artificiels - le connexionnisme - s'est largement développée depuis à peine une dizaine d'années. Le connexionnisme est maintenant largement reconnu dans le monde informatique. La formalisation de la cellule élémentaire du cerveau, et la force de son utilisation au sein d'un réseau, apportent une solution élégante à de nombreux problèmes. Cette thèse s'intéresse à la parallélisation de réseaux de neurones artificiels, particulièrement dans le domaine de la classification, un champ d'application privilégié du connexionnisme. Nous allons notamment voir qu'un réseau connexionniste peut efficacement « apprendre en parallèle ». Enfin, nous mettrons en valeur les apports mutuels du connexionnisme et des sciences cognitives.

De bonnes raisons de s'intéresser au parallélisme

La simulation de réseaux de neurones est gourmande en ressources informatiques. Pour résoudre un problème, un apprentissage sur de nombreux exemples est nécessaire. Il est également indispensable de régler des paramètres pour atteindre un bon apprentissage. Enfin, certaines applications nécessitent une réponse très rapide. Au Laboratoire de l'Informatique du Parallélisme (LIP) de l'ENS Lyon, au sein duquel les travaux présentés ici ont été développés, la charge de calculs engendrée par le connexionnisme trouve une réponse grâce aux moyens informatiques les plus puissants : les machines parallèles. Dans cette thèse, de nombreuses expériences sont menées sur des machines parallèles dites « MIMD à mémoire distribuée » ainsi que sur des réseaux locaux d'ordinateurs. Ces deux architectures sont voisines et se caractérisent par de puissants processeurs reliés par un réseau de communication.

Un classifieur incrémental

Le principal modèle neuronal étudié dans ce document est un classifieur neuronal incrémental. De nouvelles cellules sont ajoutées en fonction des besoins. Le classifieur incrémental a été mis au point peu de temps avant le début de cette thèse, à

Grenoble, au Laboratoire d'Informatique Fondamentale et d'Intelligence Artificielle (LIFIA) de l'INPG. Le programme originel était consacré à la reconnaissance de dessins au trait, notamment des chiffres manuscrits ; aussi, nous intéresserons-nous également à ce domaine d'application. Au-delà de la reconnaissance de formes, le classifieur incrémental a été utilisé sur une tâche classique de classification afin de le confronter à d'autres modèles. Le classifieur a aussi été adapté et mis en œuvre pour modéliser des tâches cognitives.

Généricité des parallélisations

La généricité du travail de parallélisation réalisé ne souffre pas de la relative marginalité du modèle étudié. Bien au contraire, nous verrons que les solutions proposées, et notamment la plus efficace, se généralisent facilement à toute une classe de réseaux de neurones artificiels performants. De plus, le modèle étant incrémental, les parallélisations proposées font face au pire cas (évolution de la charge de travail) et sont parfaitement adaptées à un réseau de neurones de taille fixe.

Les voies de la parallélisation

Plusieurs parallélisations du classifieur sont proposées et implémentées, à divers niveaux de parallélisme, et sur diverses architectures parallèles. Nous verrons comment on peut apprendre bien, tout en s'éloignant des algorithmes séquentiels généralement mis en œuvre pour simuler un réseau de neurones sur un ordinateur. Une approche modulaire efficace sera développée, nous verrons notamment en quoi des modules spécialisés et coopérants peuvent améliorer les résultats.

Nous nous fixons l'objectif de concevoir et de programmer un classifieur incrémental performant et portable (c'est-à-dire indépendant des moyens matériels et logiciels utilisés). Ce but guidera les différentes parallélisations jusqu'à obtenir une solution simple et efficace.

Une approche expérimentale

La réalisation effective de programmes opérationnels permet d'appuyer les choix théoriques par des expérimentations poussées. Une démonstration purement théorique de la validité de chaque parallélisation aurait permis de se passer de quelques expériences, mais une seule expérience peut rendre caduque un résultat théorique basé sur de mauvaises hypothèses. Nous justifions l'approche expérimentale par la difficulté de mettre en place des hypothèses assez fiables pour arriver à un résultat utilisable en-dehors d'un laboratoire. En effet, ces hypothèses se doivent de prendre en compte la complexité des moyens à mettre en œuvre lors d'une parallélisation (moyens matériels, logiciels, et humains).

Sciences cognitives

Les sciences cognitives résultent d'un désir de rapprochement entre tous les chercheurs s'intéressant à la cognition, notamment des neurologues, des psychologues, et des informaticiens. Nous utiliserons le classifieur incrémental comme un outil d'investigation pour évaluer les données collectées lors d'expériences de reconnaissance d'odeurs. Par ailleurs, le classifieur deviendra un outil de modélisation en simulant un phénomène de comportement bien connu en psychologie : « l'amorçage de répétition ». Nous profiterons de notre réflexion sur les processus de mémorisation humains pour proposer une mémoire associative hétérogène, ainsi qu'une stratégie de parallélisation.

Plan commenté

Les deux premiers chapitres introduisent les domaines de recherche auxquels ce travail appartient, l'un sur le connexionnisme (chapitre 1), l'autre sur le parallélisme (chapitre 2). Suivent deux parties distinctes, mais traitant des mêmes notions de classification, de parallélisme, et de modularité. Les chapitres 3, 4, et 5 présentent et développent différentes parallélisations du classifieur incrémental. Les chapitres 6 et 7 présentent les recherches effectuées dans le domaine des sciences cognitives.

- Le chapitre 1 de la thèse introduit le domaine de la classification. Ce premier chapitre présente le classifieur incrémental sur l'application de reconnaissance de formes pour laquelle il a été originellement conçu. Une deuxième application est étudiée afin de situer ce modèle par rapport à d'autres classifieurs couramment utilisés. À l'intention du lecteur non familier des réseaux de neurones artificiels, le chapitre est complété par l'annexe A où toutes les notions connexionnistes indispensables à la compréhension de la thèse sont présentées.
- Le chapitre 2 présente le domaine du parallélisme et son intérêt pour le connexionnisme. Cette analyse nous conduit à définir le niveau de parallélisme à mettre en œuvre : nous parlerons de « granularité ». Nous nous attarderons notamment sur l'utilisation de machines virtuelles, en l'occurrence les réseaux locaux de stations de travail. Là encore, le lecteur non spécialiste du domaine trouvera dans l'annexe B le détail de toutes les notions de base sur le parallélisme, indispensables à une bonne compréhension de la thèse.
- Le chapitre 3 présente une première parallélisation, basée sur un découpage du réseau de neurones, appelée « parallélisation par partage de l'espace d'entrée ». Une machine parallèle permet de faire évoluer ces « morceaux de réseau » simultanément. Ce travail est appliqué à la reconnaissance de chiffres manuscrits. Cette tâche de classification suppose un lourd prétraitement que nous

détaillons et parallélisons également. Une série de tests permet de mettre en lumière les avantages et les inconvénients de ce mode de parallélisation.

- Le chapitre 4 traite d’une deuxième stratégie de parallélisation qui nous occupera également au chapitre 5. Cette fois, des réseaux de neurones complets, des « modules » travaillent en même temps, toujours sur l’application de reconnaissance de formes manuscrites. Une analyse des avantages de cette approche modulaire est détaillée, notamment face à la stratégie déployée au chapitre précédent. Nous verrons par exemple que le « classifieur modulaire » résout le problème au moins aussi bien que le classifieur séquentiel, sans pour autant suivre le même comportement.
- Le chapitre 5 reprend la parallélisation modulaire, non plus sur une machine parallèle, mais sur un réseau de stations de travail. L’étude des différences entre une batterie d’ordinateurs de bureau connectés à un réseau local, et une machine dédiée au parallélisme, nous permet de réaliser une nouvelle version du classifieur modulaire. Nous verrons qu’il est plus délicat d’orchestrer le déroulement du traitement sur un réseau local de stations, que sur une machine parallèle. Aussi, notre approche est-elle basée sur une collaboration entre les machines du réseau local : nous parlerons d’approche « asynchrone ». Là encore, nous montrerons l’interaction entre parallélisme et apprentissage.
- Le chapitre 6 présente les travaux réalisés dans le cadre d’un projet de recherche du domaine des sciences cognitives. Les objectifs du projet sont présentés, ainsi que les actions menées. L’axe « traitement cognitif des odeurs » est ensuite détaillé. Nous donnons le protocole établi, et nous détaillons les expérimentations réalisées sur des sujets humains. Le classifieur incrémental est utilisé pour analyser les données collectées. L’étude du comportement du modèle nous conduit à mettre en œuvre un nouvel algorithme d’apprentissage pour le classifieur, plus efficace sur les problèmes difficiles ; cette nouvelle stratégie s’accorde avec les conclusions issues des parallélisations.
- Le dernier chapitre, le chapitre 7, couvre le deuxième axe de recherche du projet présenté au chapitre précédent : la « perception intelligente ». Comme dans le chapitre 6, l’informatique devient une des composantes des sciences cognitives. Nous montrerons comment le phénomène naturel « d’amorçage », mis en évidence par les psychologues, peut être simulé par un réseau de neurones artificiels (en l’occurrence le classifieur) par des modifications algorithmiques. Le chapitre se termine par la proposition d’un modèle connexionniste plus large, une « mémoire associative hétérogène », dans lequel le classifieur prend naturellement sa place.