

Introduction à la reconnaissance de caractères
par reseaux neuronaux

Aziez Lyes
Semeria Agostino

Coordinateur P. Collard

2002 / 2003

Chapitre 1

Introduction

1.1 Biologie

Les humains peuvent traiter beaucoup de types d'information plus rapidement que les machines. Par exemple, comprendre la langue parlée ou reconnaître des objets dans une scène visuelle - ces opérations prennent quelques cent millisecondes pour le cerveau mais beaucoup plus pour un ordinateur - si même l'opération est possible!

Mais, les unités de l'informatique de base du cerveau humain (neurones) ayez une vitesse de l'informatique de base de quelques millisecondes, beaucoup plus lente que les transistors dans un ordinateur moderne.

Caractéristiques principales:

- N entrées (quelques synapses sur chaque dendrites) et une sortie (l'axon)
- La cellule donne un signal quand la somme des entrées dépasse une valeur limite (le seuil ou 'threshold' en anglais)
- Il y a quelques rapports (synapses) plus efficaces que les autres
- Si un rapport (synapse) est fréquemment utilisé, il devient plus effectif (la force du rapport augmente)
- Le corps cellulaire principalement (soma) a une mémoire

Les signaux sont transmis entre neurones par les pulsations électriques qui voyagent le long de l'axon. Ces pulsations s'heurtent sur le neurone afférent à terminaux appelé des synapses. Ceux-ci sont trouvés sur un ensemble de pousser des branches des processus qui émergent du corps cellulaire principalement (soma) connu comme dendrites.

Chaque pulsation qui se produit à un initiés de la synapse la parution d'une petite quantité de substance chimique ou neurotransmetteur qui voyagent à travers le synaptic a fendu et lequel est reçu à récepteur de synaptic après alors place sur le dendation de pulsations. La dimension et type

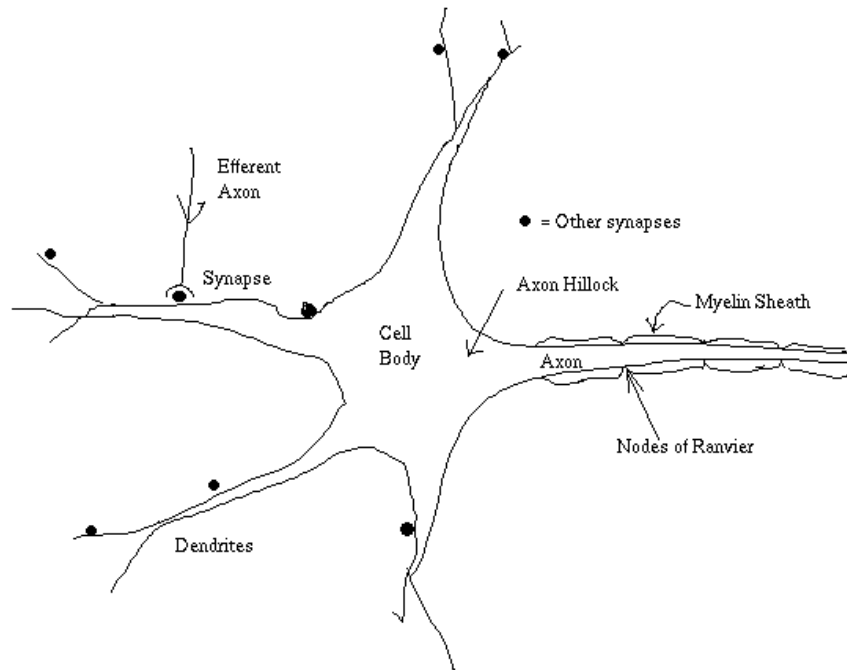


FIG. 1.1 – *Neurone humain*

de PAS produits dépendront des facteurs tel que la géométrie de la synapse et le type de neurotransmetteur.

Chaque PAS voyagera le long de son dendrite et s'est étendu sur le soma, en arrivant à la base de l'axon finalement (axon petite colline). Le neurone afférent additionne ou intègre les effets de milliers de tel PASs sur son arbre du dendritic et avec le temps. Si la capacité intégrée à l'axon petite colline dépasse un seuil, le neurone produit une action potentiel ou pointe qui commence à voyager le long de son axon. Cela commence encore la séquence entière d'événements dans neurones contenus dans le chemin de l'efferent alors.

Chapitre 2

Le reseau artificiel

2.1 Le neurone artificiel

Un neurone est une sorte d'automate, il se compose d'entrées, d'une fonction de sommation, d'une fonction de transfert (transition) et d'une ou plusieurs sorties. La figure 2.1 illustre un neurone de ce type.

Ce neurone est caractérisé par n connexions en entrée de valeurs w_i qui sont connectées à n entrées e_i ainsi que par un seuil d'activation s . L'état en sortie du neurone est déterminé par une fonction d'activation F qui renvoie -1 (neurone inactif) pour $\sum w_i x_i \leq s$ et +1 (neurone actif) dans le cas contraire. La figure 2.1 représente graphiquement cette fonction de sortie. On forme un réseau en connectant des neurones entre eux par des liens pondérés. C'est le renforcement de certaines connexions, ou leur affaiblissement, qui adapte la structure du réseau pour donner des réponses différentes. La topologie du réseau définira sa capacité d'apprentissage, sa robustesse et sa rapidité d'évaluation.

Il existe à l'heure actuelle de nombreuses architectures de réseaux de neurones. On peut les subdiviser en deux grandes catégories selon la nature de leur algorithme d'apprentissage. Les premiers sont dits "supervisés" car, lors de l'apprentissage, ils doivent disposer d'un professeur capable de leur dire ce qu'ils produisent en sortie pour chacune des données d'entraînement. Les seconds sont dits "non supervisés" car ils arrivent à s'auto-organiser.

Pour les réseaux à apprentissage supervisés (Perceptron, Adaline, etc.), on présente au réseau des entrées, et au même temps les sorties que l'on désire pour cette entrée. Par exemple on lui présente en entrée une lettre "a" manuscrite et en sortie un code correspondant à la lettre "a". Le réseau doit alors se reconfigurer, c'est-à-dire calculer ses poids afin que la sortie qu'il donne corresponde bien à la sortie désirée. Pour les réseaux à apprentissage

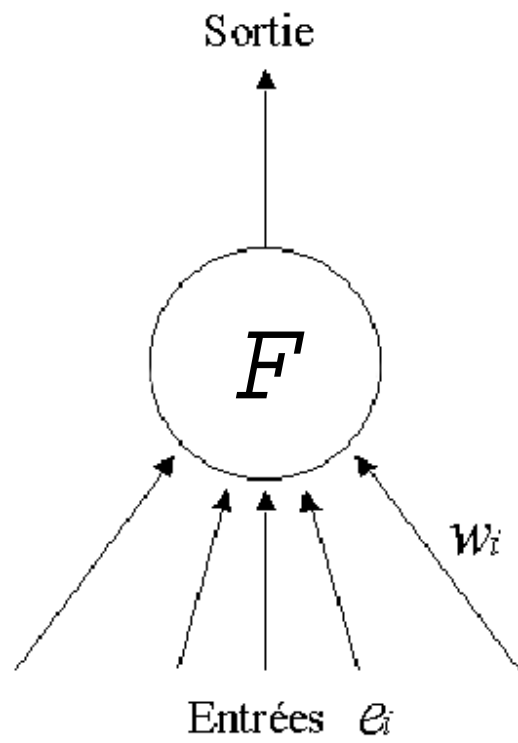


FIG. 2.1 – *Neurone artificiel*

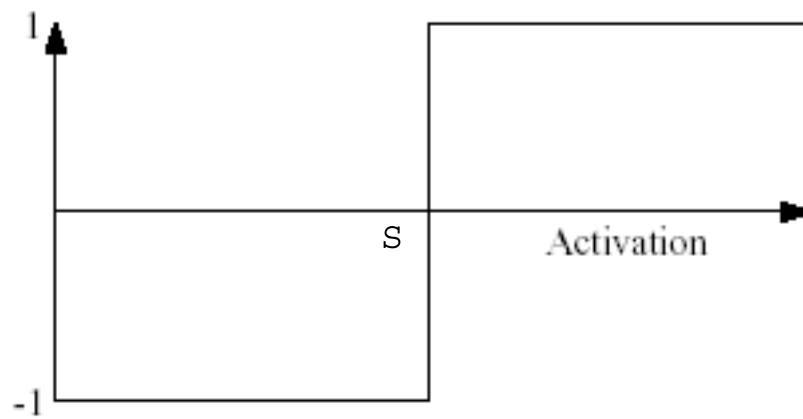


FIG. 2.2 – *la fonction de sortie d'un noeud*

non supervisé (Hopfield, Kohonen, etc.), on présente une entrée au réseau et on le laisse évoluer librement jusqu'à ce qu'il se stabilise. Il existe aussi des réseaux à apprentissage dit semi-supervisés (renforcement learning) qui ne tiennent compte que d'une évaluation partielle ou qualitative des sorties.

2.2 Vocabulaire

2.2.1 Apprentissage, mémoire et oubli

Une caractéristique des réseaux de neurones est leur capacité à apprendre (par exemple à reconnaître une lettre, un son...). Mais cette connaissance n'est pas acquise dès le départ. La plupart des réseaux de neurones apprennent par l'exemple (de la même manière qu'un enfant apprend à reconnaître un chien à partir d'exemples de chiens qu'il a vu). Ils ont donc une capacité à classer, généraliser, à mémoriser et aussi à oublier. Un réseau de neurones reconnaîtra d'autant plus facilement un objet qu'il l'aura "vu" souvent.

2.2.2 Connectivité

La connectivité des réseaux, c'est à dire la manière dont les neurones sont connectés, peut être totale (tous les neurones sont connectés entre eux) ou par couche (les neurones d'une couche ne sont connectés qu'aux couches suivantes - il existe des réseaux monocouches ou multicouches (multi layer networks)).

2.2.3 Calcul des poids synaptique

La rétropropagation (backpropagation) est une méthode de calcul des poids pour un réseau à apprentissage supervisé qui consiste à minimiser l'erreur quadratique de sortie (somme des carrés de l'erreur de chaque composante entre la sortie réelle et la sortie désirée).

D'autres méthodes de modification des poids sont plus "locales", chaque neurone modifie ses poids en fonction de l'activation ou non des neurones proches. C'est le cas des réseaux suivant la règle de Hebb (Hebbian rule) ou les réseaux de Kohonen.

2.3 Quelques réseaux célèbres

2.3.1 Les réseaux de Hopfield

Un réseau de Hopfield réalise une mémoire adressable par son contenu. Il s'agit d'un réseau constitué de neurones de McCulloch et Pitts (à deux états, -1 et 1 ou 0 et 1), dont la loi d'apprentissage est la règle de Hebb (1949), qui veut qu'une synapse améliore son activité si et seulement si l'activité de ses deux neurones est corellée (C'est à dire que le poids w_{ij} d'une connexion entre un neurone i et un neurone j augmente quand les deux neurones sont activés au même temps).

2.3.2 Les réseaux de Kohonen

Contrairement aux réseaux de Hopfield où les neurones sont modélisés de la façon la plus simple possible, on recherche ici un modèle de neurone plus proche de la réalité.

Ces réseaux sont inspirés des observations biologiques du fonctionnement des systèmes nerveux de perception des mammifères.

Une loi de Hebb modifiée (tenant compte de l'oubli) est utilisée pour l'apprentissage. La connexion est renforcée dans le cas où les neurones reliés ont une activité simultanée, et diminuée dans le cas contraire.

Une loi d'interaction latérale (observée biologiquement) est aussi modélisée. Les neurones très proches (physiquement) interagissent positivement (le poids des connexions est augmenté autour quand une synapse est activée), négativement pour les neurones un peu plus loin, et pas du tout pour les neurones éloignés. Ceci crée un "amas" de neurones activés et contribue à spécialiser certains neurones: pour une entrée donnée, une sortie particulière sera activée et pas les autres. On utilise aussi parfois des lois de concurrence entre les neurones (création et destruction de neurones selon certains critères, par exemple la règle du "winner-take-all"). Ceci permet de résoudre certains problèmes, dits NP complets, tels le problème du voyageur de commerce (comment relier n villes par le chemin le plus court). Les réseaux de Kohonen ont des applications dans la classification, le traitement de l'image, l'aide à la décision et l'optimisation.

2.3.3 Le perceptron

Une des premières tentatives de réalisation de réseaux de neurones artificiels sur ordinateur revient à Rosenblatt (1958). Dès la fin des années 50, il

en proposait une, appelée le perceptron. Ce dispositif tentait de reproduire notre capacité:

- à apprendre à associer des réponses adéquates aux stimuli auxquels nous sommes confrontés
- et à pouvoir fonctionner en présence d'informations dégradées

La conception du perceptron s'inspire explicitement du neurone.

Rosenblatt (1958) propose un algorithme permettant d'entraîner un tel réseau de neurones en ajustant les poids w_i . Il crée alors le premier réseau de neurones apprenant. La fonction d'apprentissage permet de modifier le poids des liens entre unité d'entrée et de sortie comme suit:

$$w_{nouveau} = w_{ancien} + \beta y x$$

$$\text{où } \beta = \begin{cases} +1, & \text{si la réponse est correcte} \\ -1, & \text{si la réponse est erronée} \end{cases}$$

et y est la réponse du perceptron.

De manière plus concrète, on peut imaginer, par exemple, que les différentes excitations afférentes sont liées à l'intensité lumineuse avec laquelle certains points de la "rétine" sont éclairés à partir d'un objet donné. La réponse du perceptron doit être positive ou négative suivant que l'objet est d'un certain type ou non. Donnons un exemple de mode de fonctionnement possible. Le perceptron qui nous intéresse ici doit pouvoir reconnaître une lettre "X" à un endroit donné de la "rétine". La figure 2.3 représente notre perceptron.

Chaque unité d'entrée reçoit une activation x_i - il y en a 25 dans l'exemple donné. Supposons que ces entrées correspondent à certains points situés sur la rétine. Si un point est noir, x_i prendra la valeur 1, si un point est blanc, il prendra la valeur -1. Les activations x_i vont être transmises à une unité de sortie via un lien de poids w_i . La valeur d'excitation I de l'unité de sortie sera la somme du produit des activations x_i et des poids des liens w_i . Si cette somme dépasse le seuil T , l'unité de sortie répondra positivement.

En cours d'apprentissage les poids des connexions vont être modifiés. Prenons un exemple, l'unité de sortie a répondu négativement alors que le contraire était désiré ($\beta = -1$). L'algorithme d'apprentissage va modifier une connexion dont la valeur est, par exemple -0.2 ($w_{ancien} = -0.2$), connexion liant un point d'entrée noir (dont la valeur d'activation est $x=1$) à l'unité de sortie dont la réponse est -1 ($y = -1$) comme suit :

$$w_{nouveau} = -0.2 + (-1 * (-1) * 1) = 0.8$$

Pour une connexion de même poids reliant cette fois un point blanc à l'unité de sortie, la valeur serait devenue -1.2. Après quelques essais le perceptron est sensible au stimulus et parvient à le catégoriser de manière efficace.

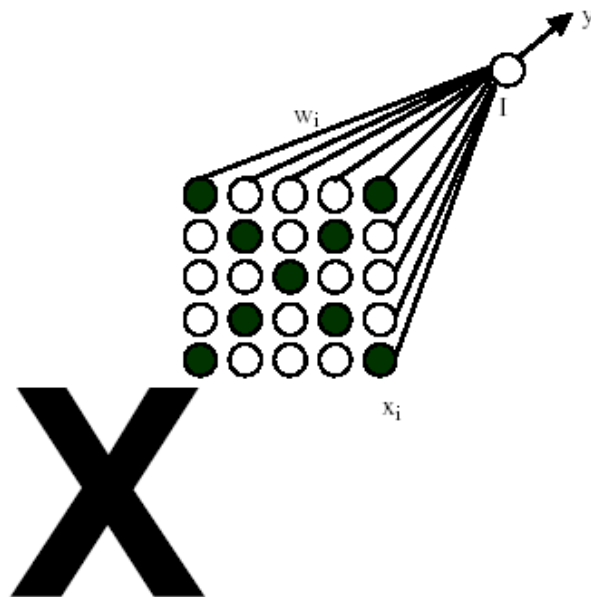


FIG. 2.3 – reconnaissance de la lettre *X* par un perceptron

Dans ce type d'architecture, la reconnaissance d'objets et l'apprentissage des associations entre objets sont primordiaux. L'association d'événements (et d'objets), plutôt que, à résolution de problèmes, est considérée comme un mécanisme de base de la cognition. L'espoir secret était d'être capable, en interconnectant un nombre suffisant d'éléments simples de ce type, de reproduire les performances du cerveau. On a réussi à montrer que les perceptrons étaient capables d'apprendre à reconnaître des formes en modifiant progressivement les poids des liens. Il était donc permis d'imaginer des systèmes qui se configurent spontanément en fonction des tâches à réaliser, les rétroactions du milieu guidant le système dans ses apprentissages. Malheureusement, une étude mathématique des perceptrons a mis en évidence leurs limites, du moins dans leur version la plus simple (Minsky et Papert, 1969). Le célèbre problème d'XOR - il n'existe aucune combinaison de w_1, w_2 , et seuil T (en général, $T = 0$) tel qu'un perceptron puisse produire la fonction suivante:

En entrée		En sortie
0	0	1
0	1	0
1	0	0
1	1	1

a ruiné les prétentions initiales. Pour répondre aux critiques de Minsky et Papert, le perceptron multi-couche (PMC) a été développé.

2.3.4 Les PMC

Le perceptron multi-couches (PMC) diffère du perceptron essentiellement par l'ajout d'une troisième couche de noeuds entre la couche d'entrée et la couche de sortie (Figure 2.4), appelée la couche cachée.

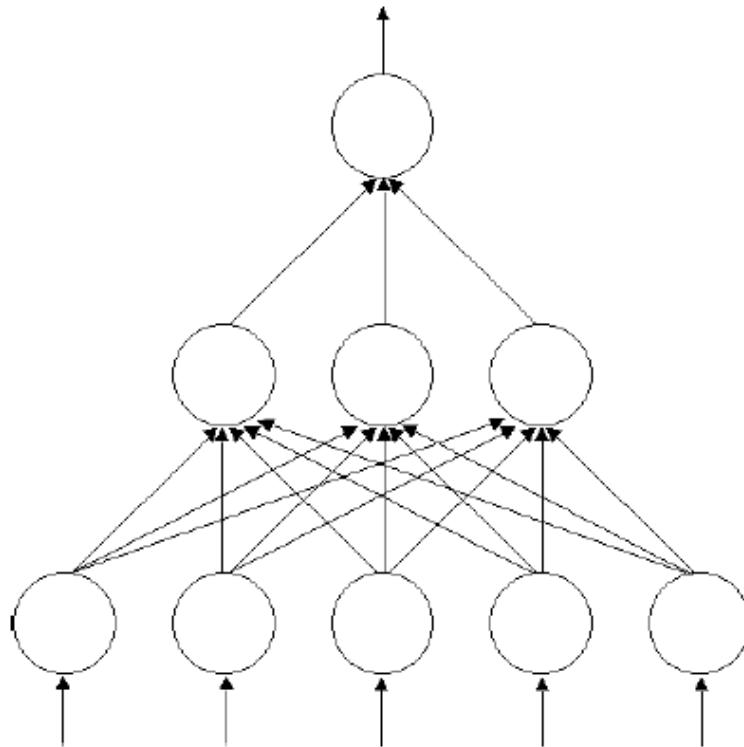


FIG. 2.4 – *Perceptron MultiCouche*

De plus, l'utilisation d'une fonction sigmoïdale (Figure 2.5) au lieu d'une simple fonction de seuil pour déterminer l'activation de sortie de chaque noeud a permis le développement d'un nouvel algorithme de changement des poids dans ces réseaux, un algorithme qui s'appelle la rétro-propagation ("backpropagation" en anglais).

Actuellement ce type de réseau et ses variantes dominent très largement les recherches en connexionnisme. Il connaît un champ d'application important, allant de la reconnaissance des visages à la conduite automatique des voitures. La difficulté d'utilisation de ce réseau réside dans le fait qu'il faut

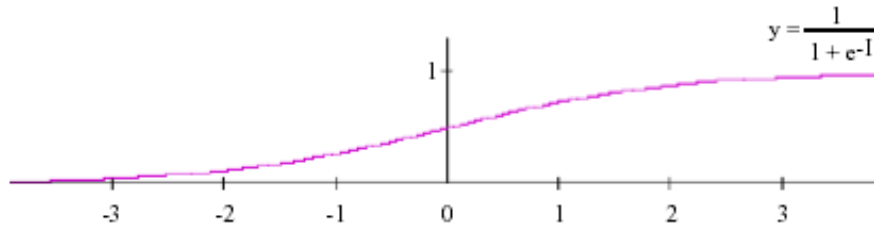


FIG. 2.5 – la fonction sigmoïdale

déterminer la topologie du réseau. Il s'agit de définir le nombre de neurones des différentes couches ainsi que leurs interconnexions. Si le nombre de neurones cachés est trop faible, l'algorithme d'apprentissage n'arrivera pas à construire une représentation intermédiaire du problème qui soit linéairement séparable, et certains des exemples ne seront pas appris correctement. Inversement si ce nombre est trop élevé, il y a risque d'apprentissage par coeur du problème. Le réseau apprend parfaitement les exemples d'apprentissage, mais ne donne que des résultats médiocres sur des nouvelles données qu'il n'a pas vues durant l'apprentissage. Une solution pour résoudre le problème du choix du nombre de neurones est d'ajouter incrémentalement des neurones jusqu'à ce que le réseau soit capable d'apprendre le problème.

Chapitre 3

La reconnaissance de caractères

La reconnaissance de caractères manuscrits revêt une importance particulière, du fait que comme la reconnaissance de la parole, elle s'inscrit dans un schéma général de communication hommemachine.

On peut très bien imaginer des systèmes fiables, permettant de donner des ordres à une machine au moyen de l'écrit. De tels systèmes seraient aussi capables d'intégrer des bases de données manuscrites, de reconnaître un code postal, une somme sur un chèque bancaire, voire pourquoi pas, de reconnaître le scripteur. Selon le type d'application envisagé, on peut opérer plusieurs distinctions sur l'approche à employer ainsi que sur les performances à obtenir.

Les techniques dites online, utilisent un capteur spécifique (tablette graphique, stylo électronique) pour acquérir les informations. Ces informations reflètent la dynamique de l'écriture, dynamique pouvant se révéler fort utile pour segmenter des caractères (i.e. isoler chaque caractère), voire pour le reconnaître explicitement (ainsi l'ordre du dessin d'un caractère chinois influence son sens).

Les techniques dites offline, permettent de travailler sur une image fournie par un scanner. Cette approche est adaptée au traitement de documents existants (chèques, textes, ...). Par contre le manque d'information dynamique y rend la segmentation plus difficile.

On pourra tolérer des performances moyennes dans le cas de la reconnaissance de texte. Les erreurs pourront en effet ensuite être traitées à l'aide d'un dictionnaire. Dans le cas de l'identification d'un scripteur (vérification d'une signature, autorisation d'accès, ...), de hautes performances sont par contre nécessaires.

Les applications de notre travail sont plutôt à envisager dans le cadre d'un système offline nécessitant de bonnes performances.

3.1 Le système OCR

Il consiste en général en un ensemble de modules de prétraitement, segmentation, extraction de traits, classification et traitement contextuel. Un document papier est dans un premier temps numérisé pour obtenir une image en niveaux de gris (éventuellement en noir et blanc).

Durant le prétraitement, on filtre le signal pour éliminer le bruit, les zones contenant du texte sont détectées et converties en images binaires à l'aide d'une méthode de seuillage, soit globale, soit localement adaptative. Durant la phase de segmentation, les images texte sont séparées en caractères distincts. Ce qui est extrêmement délicat dans le cadre de textes manuscrits du fait de la connexité des caractères. Une des techniques souvent mise en uvre consiste à découper les patrons en unités plus petites (sur-segmentation), puis à trouver les points de segmentation corrects à l'aide des sorties d'un classificateur de patrons. Du fait de la grande variation des écritures et d'un haut niveau de bruit, reconnaître des caractères segmentés n'est pas chose aisée. Le lecteur pourra en juger sur la figure 3.1 qui montrent des exemplaires normalisés en taille d'un échantillon de la base de caractères manuscrits du National Institute of Standards and Technology (NIST).

La figure 3.2 présente deux manières d'utiliser un RNA (reseau neuronal artificiel) dans un système OCR. La première utilise un extracteur de traits (pas nécessairement de type RNA). Par exemple, les traits caractéristiques peuvent être les contours des lettres, qui sont donnés en entrée au réseau non bouclé. Cette approche est extrêmement flexible puisqu'elle permet d'incorporer une grande variété de caractéristiques. L'autre approche n'utilise pas de manière explicite une extraction des traits à partir des données brutes. L'extraction se fait de façon implicite dans les couches cachées du réseau. L'avantage de cette méthode est que l'extraction et la classification sont intégrées et entraînées simultanément afin de produire une classification optimale. Il n'est pas établi que l'extraction implicite soit plus efficace dans le cadre de la reconnaissance de caractères et qui plus est, cette approche nécessite un réseau de taille beaucoup plus conséquente que la première méthode.

Le Cun a mis en oeuvre cette seconde technique pour la reconnaissance de codes postaux(LeNet). On présente en entrée au réseau une image normalisée en niveaux de gris à un réseau possédant trois couches cachées. La première couche réalise une cartographie des traits localisés. Les cellules de la seconde couche combinent les informations de la première couche. Le niveau d'activation en sortie pouvant s'interpréter comme une approximation de la probabilité à posteriori qu'un caractère appartienne à une classe donnée. Les catégories en sortie sont ordonnées en fonction des niveaux d'activation et sont fournies à l'étape de post-processing. À cette étape, l'informa-

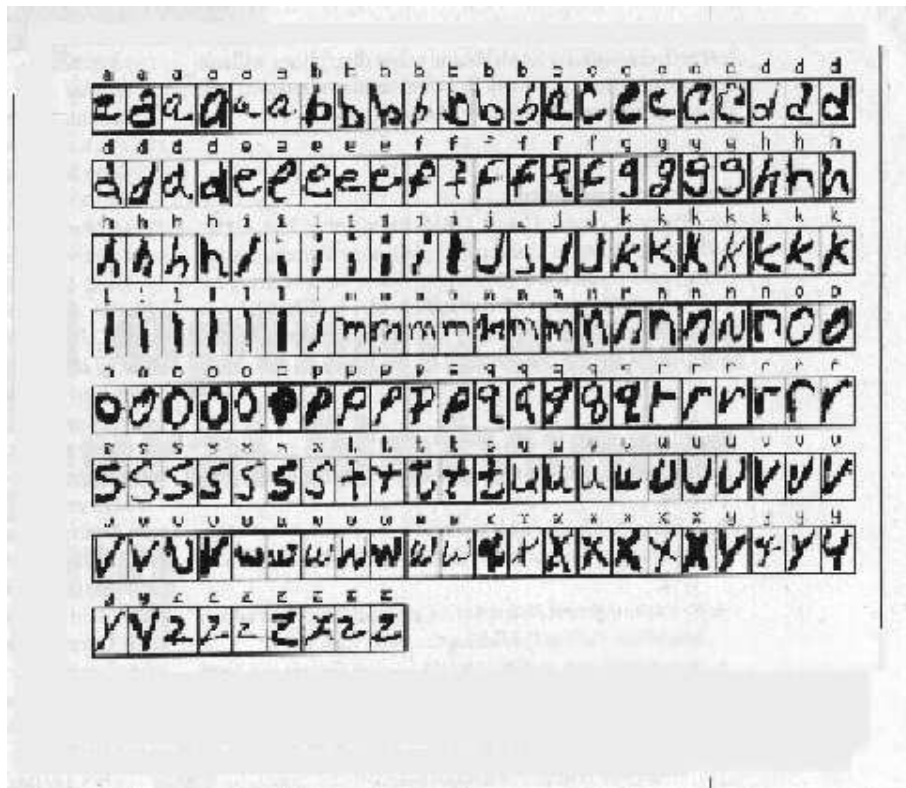


FIG. 3.1 – Échantillon de caractères manuscrits dans la base du NIST

tion contextuelle est utilisée pour mettre à jour les classifications. Ceci peut mettre en jeu, par exemple, un dictionnaire des mots admissibles ou utiliser des contraintes syntaxiques présentes dans un numéro de téléphone ou de sécurité sociale.

Les RNA sont performants dans le cas des systèmes OCR. Cependant, aucun résultat ne permet d'établir leur supériorité par rapport à des techniques classiques comme les classificateurs statistiques. Lors de la première conférence en 1992 "Census Optical Character Recognition System Conference", plus de quarante systèmes de reconnaissance de caractères manuscrits furent évalués sur une base commune. Les dix premiers utilisaient soit un type particulier de réseaux non bouclés, soit un classificateur basé sur les plus proches voisins.

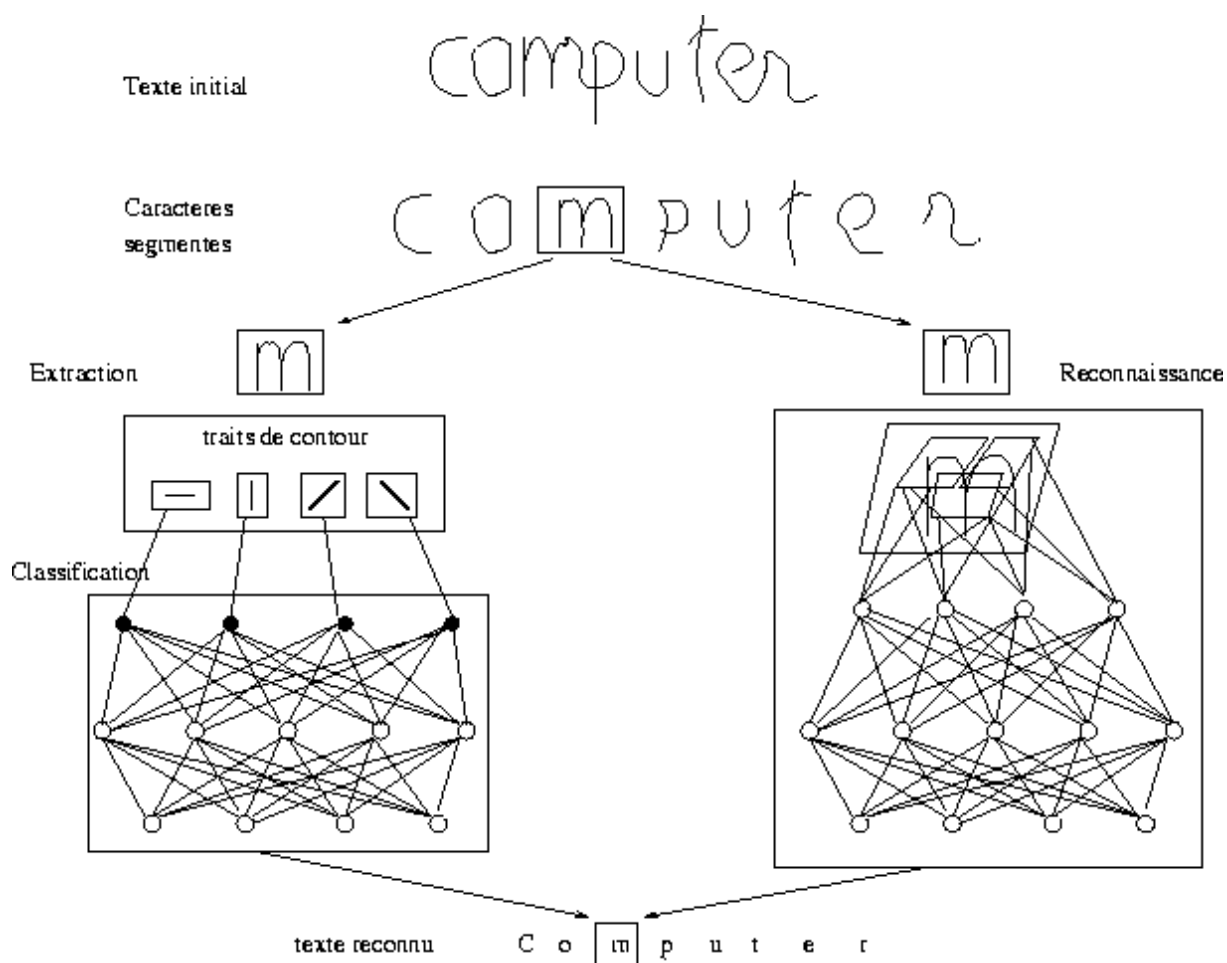


FIG. 3.2 – Deux réseaux dans le cadre des OCR

3.2 LeNet

Ce type d'architecture a été proposé par Le Cun en 1989 et progressivement améliorée. C'est une architecture de type perceptron multicouches à 4 couches cachées.

Les données de taille 16 sont présentées sur une rétine de taille 28 x 28.

La bordure ainsi ajoutée évite des effets de bords pouvant apparaître de par l'utilisation de masques sur les connexions vers la première couche cachée.

Les connexions vers la seconde couche cachée sont de types poids partagés et calculent une moyenne locale. Le couple premièredeuxième couches cachées effectue ainsi une extraction de caractéristiques suivie d'un sous-échantillonnage.

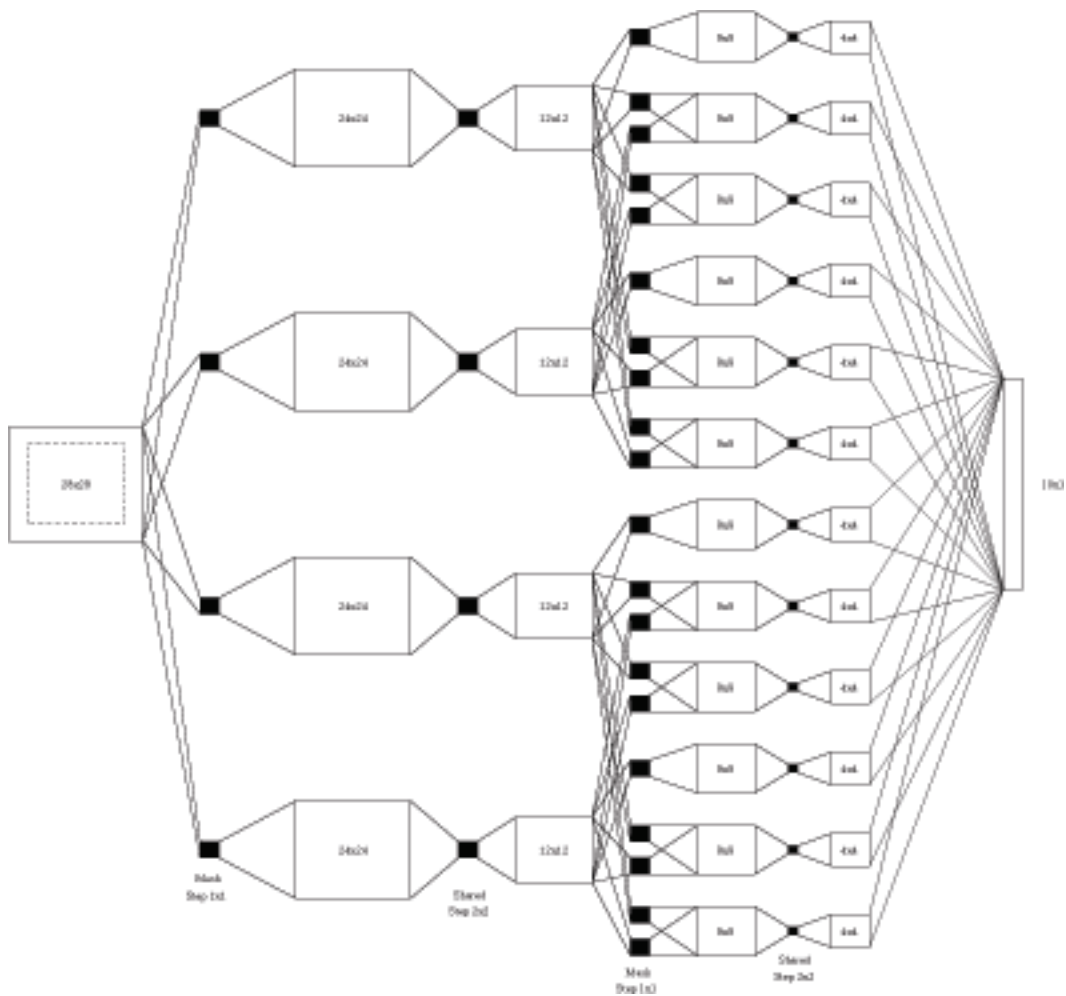


FIG. 3.3 – Architecture du réseau LeNet

Les deux couches suivantes effectuent le même travail (extraction puis souséchantillonnage). Enfin, des connexions totales sont utilisées vers la dernière couche, ce qui permet de réaliser une classification quasilineaire sur les caractéristiques extraites dans l'avant dernière couche.

Au total, LeNet comporte 4634 cellules, 98442 connexions et 2578 paramètres libres (du fait de l'utilisation de poids partagés). Ce qui rend le nombre de calculs à effectuer assez conséquents.

Les performances sont par contre excellentes, et LeNet reste encore aujourd'hui, l'architecture connexionniste la plus performante.

Chapitre 4

Conclusion

4.1 Conclusion

Les RNA étaient plus efficaces en terme de rapidité et d'occupation mémoire que les classificateurs. À la différence de ces derniers, la vitesse des RNA était indépendante de la taille de la base de données choisie. Le taux de reconnaissance pour ces meilleurs systèmes OCR sur les données du NIST (caractères isolés et pré-segmentés) était au delà de 98La différence constatée sur les minuscules est largement liée au fait que les données tests différaient grandement des données d'apprentissage et dans un même temps à des erreurs plus basiques.

Il ressort de ces tests que les systèmes OCR ont des performances comparables aux performances humaines dans le cas de la reconnaissance de caractères isolés, mais que l'analyse humaine est encore sans égale pour les documents manuscrits sans contrainte typographique.

Les récents développements dans le domaine des RNA ont suscité à la fois enthousiasme et critiques. Pour de nombreuses tâches, telles que la reconnaissance de formes, aucune approche n'est vraiment supérieure.

Le choix de la meilleure technique doit en fait être guidé par la nature du problème à résoudre. Il faut chercher à mieux comprendre les capacités, les hypothèses d'utilisation et les domaines d'applications des différentes approches afin d'utiliser au mieux leur complémentarité en vue d'obtenir des systèmes plus intelligents.

Il semble qu'une approche de ce type permettrait une synergie dont l'intérêt serait de combiner les avantages des différentes techniques à même de résoudre un problème. Comme Mynsky l'a récemment observé, il est temps de combiner des systèmes multi-composantes, qui nécessitent une bonne technologie d'intégration.

Il est clair que la coopération et la communication entre chercheurs du domaine des RNA et d'autres disciplines permettra/ permettrait d'éviter des travaux redondants mais, et plus important encore, stimulera et bénéficiera à chacune des disciplines

Bibliographie

- [1] Le Cun. Back-propagation applied to handwritten zipcode recognition. *Neural Computation*, 1:-551, 1989.
- [2] Wilkinson. The first census optical character recognition system conference. Tech. Report NISTIR 4912, US Dept. Commerce, NIST, Gaithersburg, Md., 1992.
- [3] Le cerveau et la pensée edition Science
- [4] Olivier Grzelak "Apprentissagesupervise et extraction de connaissance à partir des reseaux de neurones de types hyperconvexe"
- [5] These de Philippe Leray "Quelques types de réseaux de neurones"
- [6] Daniel Defays, Robert M. French Jacques Sougné "Apports de l'Intelligence Artificielle à la Psychologie"
- [7] <http://www.scico.u-bordeaux2.fr/corsini/Pedagogie/ANN/main/node14.html>
- [8] <http://www.grappa.univ-lille3.fr/polys/apprentissage/sortie005.html>
- [9] <http://zoid.org/tforcdivad/reseauxNeuronaux/classe1.html>

Table des matières

1	Introduction	1
1.1	Biologie	1
2	Le reseau artificiel	3
2.1	Le neurone artificiel	3
2.2	Vocabulaire	5
2.2.1	Apprentissage, mémoire et oubli	5
2.2.2	Connectivité	5
2.2.3	Calcul des poids synaptique	5
2.3	Quelques réseaux célèbres	6
2.3.1	Les réseaux de Hopfield	6
2.3.2	Les réseaux de Kohonen	6
2.3.3	Le perceptron	6
2.3.4	Les PMC	9
3	La reconnaissance de caractères	11
3.1	Le système OCR	12
3.2	LeNet	14
4	Conclusion	16
4.1	Conclusion	16
	Bibliographie	18