



Cours 14 - Réseaux de Neurones Artificiels

MARIE-HÉLÈNE ABEL

Table des matières

I - Préambule	5
II - Un peu d'histoire	7
A. Origine.....	7
B. Premiers succès.....	8
C. Frein : les limitations.....	8
D. Redémarrage.....	8
E. Levée des limitations.....	9
III - Fondements biologiques	11
A. Le cerveau.....	11
B. Le neurone.....	12
C. La synapse.....	12
IV - Le neurone formel	13
A. Modèle du neurone formel.....	13
B. Modélisation des synapses.....	14
C. Synthèse.....	15
V - Classification des Réseaux de Neurones Artificiels	17
A. Définition.....	17
1. Le type de neurone.....	17
2. L'architecture du réseau.....	18
3. La dynamique du réseau.....	23
4. Le mode d'apprentissage.....	23
B. Processus connexioniste.....	25
1. Etape 1 : la base d'exemples.....	25
2. Etape 2 : initialisation du RNA.....	26
3. Etape 3 : apprentissage.....	26
4. Etape 4 : test.....	26
5. Etape 5 : utilisation.....	27
C. Propriétés remarquables des RNA.....	27
1. La capacité de généralisation.....	27

2. <i>La rapidité</i>	27
3. <i>Une mémoire distribuée</i>	27
4. <i>La résistance au bruit</i>	28
5. <i>La résistance à la dégradation</i>	28
6. <i>L'adaptabilité</i>	28
D. Domaine d'application.....	28

VI - Le perceptron

31

Préambule



- Les réseaux de neurones formels sont à l'origine une tentative de modélisation mathématique du cerveau humain.
- Les premiers travaux datent de 1943 et sont l'oeuvre de **MM. Mac Culloch et Pitts**.
- Ils présentent un modèle assez simple pour les neurones et explorent les possibilités de ce modèle.

Idée

- On se donne une unité simple, un neurone, qui est capable de réaliser quelques calculs élémentaires.
- On relie ensuite entre elles un nombre important de ces unités.
- Et on essaye de déterminer la puissance de calcul du réseau ainsi obtenu.



Remarque

Il est important de noter que ces neurones manipulent des données **numériques** et non des données symboliques.

Un peu d'histoire



Origine	7
Premiers succès	8
Frein : les limitations	8
Redémarrage	8
Levée des limitations	9

A. Origine

1890

W. James, psychologue américain introduit le concept de mémoire associative, et propose ce qui deviendra une loi de fonctionnement pour l'apprentissage sur les réseaux de neurones connue plus tard sous le nom de loi de Hebb.

1943

J. Mc Culloch et W. Pitts introduisent une modélisation du neurone biologique (un neurone au comportement binaire).



Remarque

Ceux sont les premiers à montrer que des réseaux de neurones formels simples peuvent réutiliser des fonctions logiques, arithmétiques et symboliques complexes (tout au moins au niveau théorique).

1949

D. Hebb, physiologiste américain explique le conditionnement chez l'animal par les propriétés des neurones eux-mêmes.



Méthode

Un conditionnement de type pavlovien tel que, nourrir tous les jours à la même heure un chien, entraîne chez cet animal la sécrétion de salive à cette heure précise même en l'absence de nourriture.



Remarque

La loi de modification des propriétés des connexions entre neurones qu'il propose explique en partie ce type de résultats expérimentaux.

B. Premiers succès

1957

- F. Rosenblatt développe le modèle du Perceptron.
- Il construit le premier neuro-ordinateur basé sur ce modèle et l'applique au domaine de la reconnaissance de formes.

1960

- B. Widrow, un automaticien, développe le modèle Adaline (Adaptative Linear Element).
- Dans sa structure, le modèle ressemble au Perceptron, cependant la loi d'apprentissage est différente.
Celle-ci est à l'origine de l'algorithme de rétropropagation de gradient très utilisée aujourd'hui avec les Perceptrons multicouches.



Remarque

Les réseaux de type Adaline restent utilisés de nos jours pour certaines applications particulières.

C. Frein : les limitations

1969

M. Minsky et S. Papert publient un ouvrage qui met en exergue les limitations théoriques du perceptron.



Remarque

Limitations alors connues, notamment concernant l'impossibilité de traiter par ce modèle des problèmes non linéaires.

- Ils étendent implicitement ces limitations à tous les modèles de réseaux de neurones artificiels.
- Leur objectif est atteint, il y a abandon financier des recherches dans le domaine (surtout aux U.S.A.), les chercheurs se tournent principalement vers l'IA symbolique et les systèmes à bases de règles.

D. Redémarrage

1982

- Parution de l'article de J. J. Hopfield, un physicien reconnu.
- Présentation d'une théorie du fonctionnement et des possibilités des réseaux de neurones.



Méthode

- J. J. Hopfield fixe préalablement le comportement à atteindre pour son modèle et à partir de là, la structure et la loi d'apprentissage correspondant au résultat escompté.
- Ce modèle est aujourd'hui encore très utilisé pour des problèmes d'optimisation.



Remarque

- Notons qu'à cette date, l'IA symbolique est l'objet d'une certaine désillusion, elle n'a pas répondu à toutes les attentes et s'est même heurtée à de sérieuses limitations.
- Aussi, bien que les limitations du Perceptron mises en avant par M. Minsky ne soient pas levées par le modèle d'Hopfield, les recherches sont relancées.

E. Levée des limitations

1983

La Machine de Boltzmann est le premier modèle connu apte à traiter de manière satisfaisante les limitations recensées dans le cas du perceptron.



Remarque

L'utilisation pratique s'avère difficile, la convergence de l'algorithme est extrêmement longue (les temps de calcul sont considérables).

1985

- La rétropropagation de gradient apparaît.
- C'est un algorithme d'apprentissage adapté aux réseaux de neurones multicouches (aussi appelés Perceptrons multicouches).

Fondements biologiques



Le cerveau	11
Le neurone	12
La synapse	12

A. Le cerveau

Dans le cerveau, les neurones sont reliés entre eux par l'intermédiaire d'axones et de dendrites.

En première approche, on peut considérer que ces sortes de filaments sont conducteurs d'électricité et peuvent ainsi véhiculer des messages depuis un neurone vers un autre.



Remarque

Les dendrites représentent les entrées du neurone et son axone sa sortie.

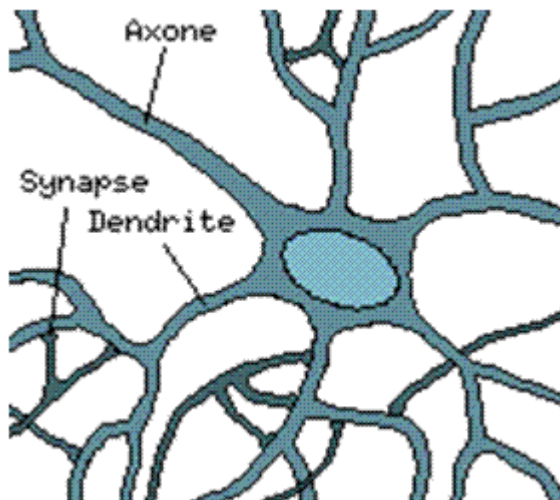


image1

B. Le neurone



Méthode

Un neurone émet un signal en fonction des signaux qui lui proviennent des autres neurones.

On observe en fait au niveau d'un neurone, une intégration des signaux reçus au cours du temps, c'est-à-dire une sorte de sommations des signaux.



Remarque

En général, quand la somme dépasse un certain seuil, le neurone émet à son tour un signal électrique.

C. La synapse

La notion de synapse explique la transmission des signaux entre un axone et une dendrite.

- Au niveau de la jonction (c'est-à-dire de la synapse), il existe un espace vide à travers lequel le signal électrique ne peut pas se propager.
- La transmission se fait alors par l'intermédiaire de substances chimiques, les neuro-médiateurs.



Méthode

- Quand un signal arrive au niveau de la synapse, il provoque l'émission de neuro-médiateurs qui vont se fixer sur des récepteurs de l'autre côté de l'espace inter-synaptique.
- Quand suffisamment de molécules se sont fixées, un signal électrique est émis de l'autre côté et on a donc une transmission.



Remarque

- En fait, suivant le type de la synapse, l'activité d'un neurone peut renforcer ou diminuer l'activité de ces voisins.
- On parle ainsi de synapse **excitatrice** ou **inhibitrice**.

Le neurone formel

IV

Modèle du neurone formel	13
Modélisation des synapses	14
Synthèse	15

A. Modèle du neurone formel

Le modèle de neurone formel dû à Mac Culloch et Pitts, est un modèle mathématique très simple dérivé d'une analyse de la réalité biologique.



Remarque

Le modèle biologique fait intervenir une notion temporelle qui est difficile à intégrer dans un modèle simple.



Méthode

- On oublie cette notion : on remplace l'intégration temporelle par une simple sommation des signaux arrivant au neurone (ces signaux sont communément appelés les entrées du neurone).
- On compare ensuite la somme obtenue à un seuil et on déduit de la comparaison la sortie du neurone.
- Cette sortie sera par exemple égale à 1 si la somme est supérieure au seuil et à 0 (ou -1) dans le cas contraire.



Méthode : Plus formellement

- Il suffit pour obtenir ce comportement de soustraire le seuil considéré à la somme des entrées, et de faire passer le résultat par la fonction de transfert du neurone qui est ici la fonction de **Heaviside**.
- Le résultat après transfert est alors la sortie du neurone.



Remarque : Cette enchaînement "sommation" puis "non-linéarité" représente finalement les propriétés "physiques" du neurone.

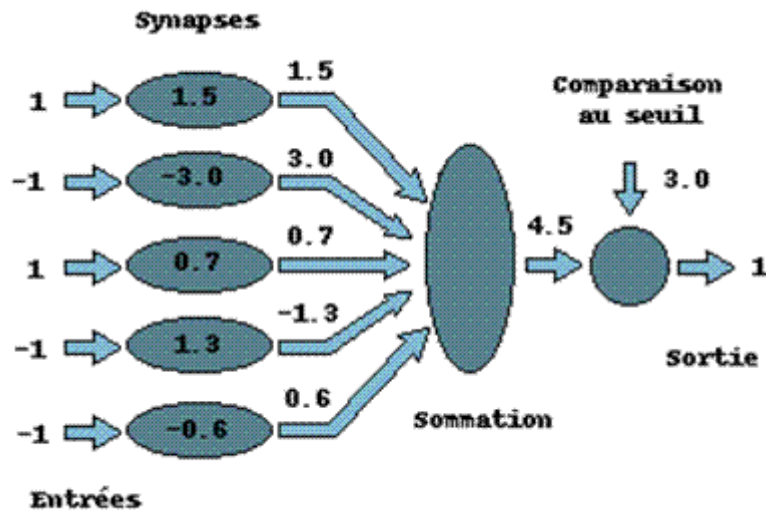


image2

B. Modélisation des synapses



Méthode : La modélisation des synapses est assez simpliste en première approche

On se contente en effet d'introduire la notion de synapse excitatrice et de synapse inhibitrice en multipliant la sortie d'un neurone par 1 ou -1 avant de la transmettre aux autres neurones.



Remarque : Afin de donner plus de puissance au modèle

- On généralise ce modèle en introduisant ensuite une connexion synaptique qui est une valeur réelle.
- La sortie d'un neurone est alors multipliée par la valeur de la connexion synaptique avant d'être transmise à un autre neurone.
- On modélise ainsi la connexion entre deux neurones.

C. Synthèse

Un neurone formel réalise :

- Une somme pondérée de ses entrées.
- Ajoute un seuil à cette somme.
- Fait passer le résultat par une fonction de transfert pour obtenir sa sortie.

Classification des Réseaux de Neurones Artificiels

V

Définition	17
Processus connexioniste	25
Propriétés remarquables des RNA	27
Domaine d'application	28

A. Définition



Définition

- Les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle.
- Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit.
- Toute structure hiérarchique de réseaux est évidemment un réseau.
- Un réseau de neurones formels est constitué d'un grand nombre de cellules de base interconnectées.
- De nombreuses variantes sont définies selon le choix de :
 - La cellule élémentaire (type du neurone).
 - L'architecture du réseau.
 - La dynamique du réseau.
 - Le mode d'apprentissage.

1. Le type de neurone

Automates à seuil

Ce sont les plus fréquemment utilisés.



Méthode : Leur principe est très simple :

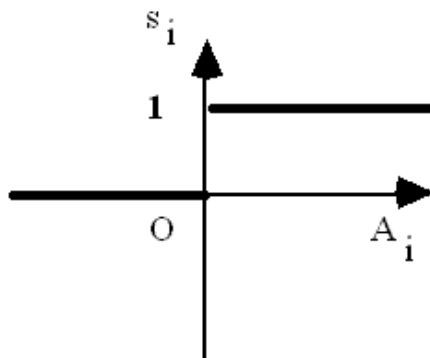
- Les états internes ne peuvent prendre que les valeurs 0 ou 1.
- Si la somme pondérée A_i des entrées d'un neurone dépasse un seuil fixé, l'état interne et la sortie de ce même neurone prennent la valeur 1 (0 dans le cas contraire).

Automates continus

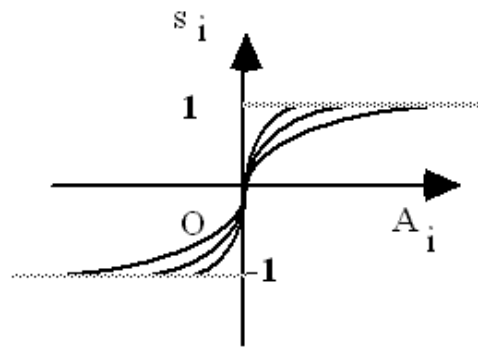
L'ensemble des valeurs des états internes est continu (valeurs réelles), la fonction de transition calculant la sortie de façon déterministe.



Exemple : Automates à seuil et continu



seuil



sigmoïde

image3

2. L'architecture du réseau

Il existe différentes topologies de réseaux correspondant aux divers schémas géométriques de connexion entre les neurones.

- Les connexions entre les neurones qui composent le réseau décrivent la topologie du modèle.
- La topologie peut être quelconque, mais le plus souvent il est possible de distinguer une certaine régularité.

a) Réseau multicouche

- Les neurones sont arrangés par couche.
- Il n'y a pas de connexion entre neurones d'une même couche.



Méthode

Habituellement, chaque neurone d'une couche est connecté à tous les neurones de la couche suivante et celle-ci seulement.



Remarque

- Notion de sens de parcours de l'information au sein d'un réseau.
- Notion de neurones d'entrée et neurones de sortie.



Définition : On appelle :

- Couche d'entrée : l'ensemble des neurones d'entrée.
- Couche de sortie : l'ensemble des neurones de sortie.
- Couches cachées : l'ensemble des couches intermédiaires, elles n'ont aucun contact avec l'extérieur.



Exemple

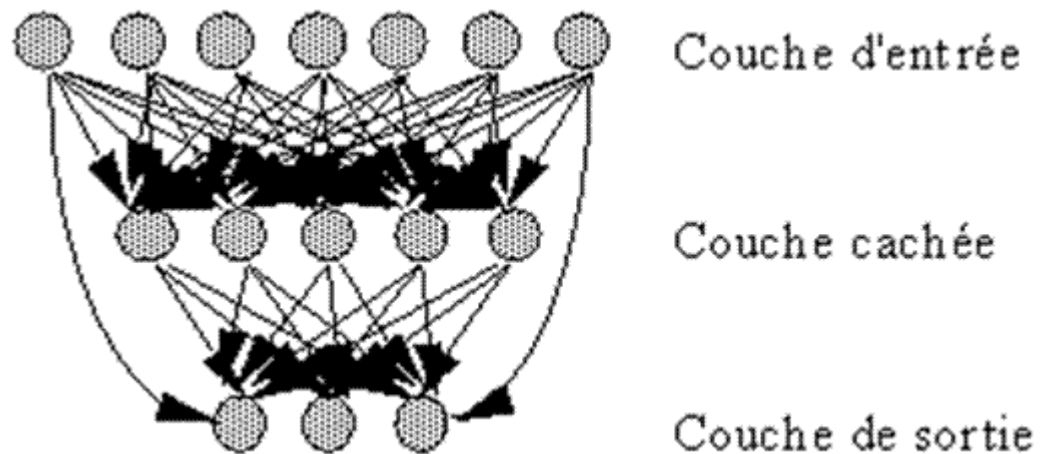


image4

b) Réseau à connexions locales



Définition

- Il s'agit d'une structure multicouche.
- Chaque neurone entretient des relations avec un nombre réduit et localisé de neurones de la couche avale.



Remarque

Les connexions sont donc moins nombreuses que dans le cas d'un réseau multicouche classique.



Exemple

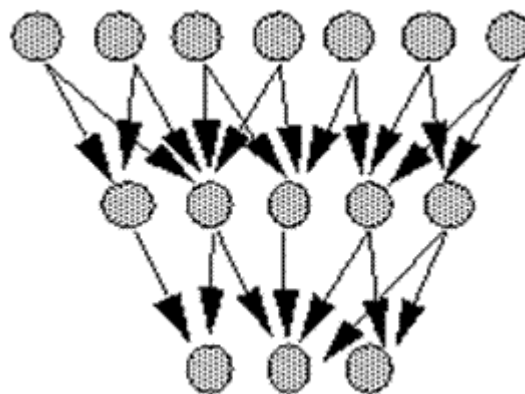


image5

c) Réseau à connexions récurrentes



Définition

- Les connexions récurrentes ramènent l'information en arrière par rapport au sens de propagation défini dans un réseau multicouche.
- Ces connexions sont le plus souvent locales.



Exemple

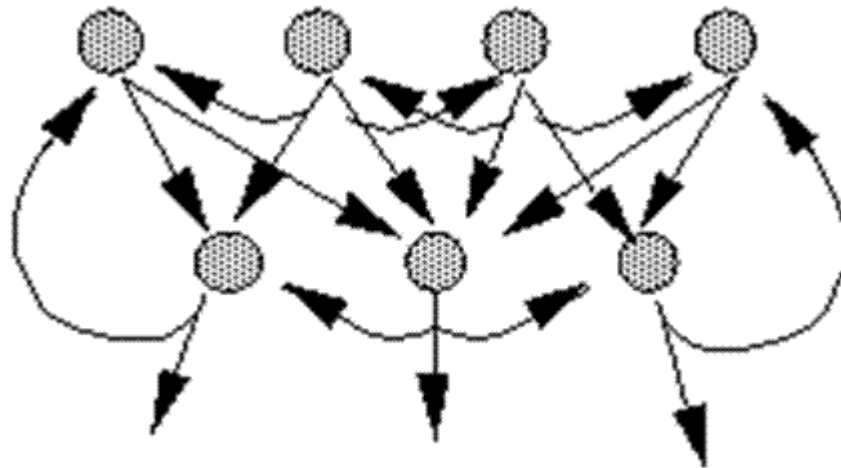


image6

d) Réseau à connexion complète

Définition

- C'est la structure d'interconnexion la plus générale.
- Chaque neurone est connecté à tous les neurones (et lui-même).

Exemple

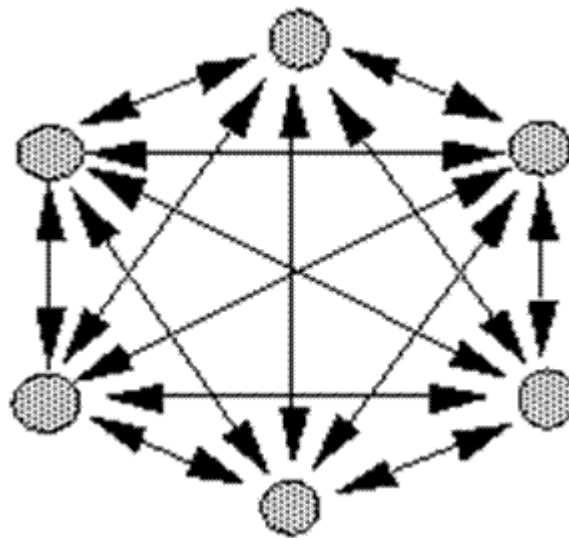


image7

3. La dynamique du réseau

Définition

La dynamique du réseau désigne l'évolution de l'état interne des neurones au cours du temps.

Elle peut être :

- Synchrones : toutes les cellules calculent leurs sorties respectives simultanément.
- Asynchrone :
 - Séquentielle : les cellules calculent leurs sorties chacune à son tour en séquence.

- Aléatoire.

A partir d'un état initial, la dynamique du réseau conduit le réseau de neurones à évoluer :

- vers des cycles,
- ou vers des états stables, encore appelés états attracteurs (par similitude avec la physique), et dont le nombre est inférieur à celui des états potentiels du réseau.



Remarque

Placé dans un état voisin d'un de ces états attracteurs, le réseau tend naturellement vers cet état de la même façon qu'une bille, lancée à proximité d'une cavité creusée dans une nappe, serait irrésistiblement attirée par et vers elle.

4. Le mode d'apprentissage

Les principales applications des réseaux de neurones sont l'optimisation et l'apprentissage.



Remarque

- L'apprentissage est vraisemblablement la propriété la plus intéressante des réseaux neuronaux.
- Elle concerne les modèles les plus utilisés.



Définition

L'apprentissage est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré.



Remarque

L'apprentissage neuronal fait appel à des exemples de comportement.



Remarque

- Dans le cas des réseaux de neurones artificiels, on ajoute souvent à la description du modèle l'algorithme d'apprentissage.
- Le modèle sans apprentissage présente en effet peu d'intérêt.

Dans la majorité des algorithmes actuels, les variables modifiées pendant l'apprentissage sont les poids des connexions.

=> L'apprentissage est la modification des poids du réseau dans l'optique d'accorder la réponse du réseau aux exemples et l'expérience.



Remarque

- Il est souvent impossible de décider a priori des valeurs des poids des connexions d'un réseau pour une application donnée.
- A l'issue de l'apprentissage, les poids sont fixés : c'est alors la phase d'utilisation.



Définition : Algorithmes d'apprentissage

Selon la forme des exemples d'apprentissage, deux grandes classes d'algorithmes sont distinguées :

- Apprentissage supervisé : les exemples sont des couples (Entrée Sortie).
- Apprentissage non supervisé : les exemples sont limités aux valeurs des entrées.



Définition : Algorithmes d'apprentissage : autre distinction

- Apprentissage << off-line >> :
Toutes les données sont dans une base d'exemples d'apprentissage qui sont traités simultanément.
- Apprentissage << on-line >> :
Les exemples sont présentés les uns après les autres au fur et à mesure de leur disponibilité.
- L'apprentissage à l'aide de réseaux de neurones est tolérant au bruit et aux erreurs.
- Le temps d'apprentissage peut être long, par contre, après apprentissage, le calcul des sorties à partir d'un vecteur d'entrée est rapide.



Attention : Le résultat de l'apprentissage (le réseau de neurones calculé par l'algorithme d'apprentissage), n'est pas interprétable par l'utilisateur :

- On ne peut pas donner d'explication au calcul d'une sortie sur un vecteur d'entrée.
- On parle de << boîte noire >>.

B. Processus connexionniste



Définition : Le processus connexionniste peut se définir en 5 étapes :

- Etape 1 : la base d'exemples
- Etape 2 : initialisation du RNA
- Etape 3 : apprentissage
- Etape 4 : test
- Etape 5 : utilisation

1. Etape 1 : la base d'exemples

Constitution d'une base d'exemples, nombreux.

- Les exemples sont des cas particuliers et représentatifs de résolution du problème à traiter ;
- Ils sont choisis de façon à ce que chaque facette du problème soit abordée.



Remarque

Il est préférable que la base d'exemples couvre et "maille" efficacement l'espace représentant les différentes configurations de données possibles.

2. Etape 2 : initialisation du RNA



Méthode

- Choix d'un RNA ayant des caractéristiques (topologie, modes opératoires, ...) adaptées au problème.
Celles-ci sont choisies heuristiquement.
- Ce RNA est ensuite initialisé à vide :
 - Tous les poids synaptiques sont aléatoires,
 - Le RNA ne sait rien !

3. Etape 3 : apprentissage



Méthode

- Le RNA extrait directement de la base d'exemples l'expertise qui y est contenue.
- Il s'auto configure, via un algorithme d'apprentissage, de façon à pouvoir résoudre chaque exemple proposé.
- Cette transformation passe par l'adaptation progressive des poids synaptiques en application d'un processus en quatre étapes :

Processus : étape 3.1

- Les entrées d'un exemple sont injectées dans les cellules d'entrée du RNA.
- Le RNA calcule alors une sortie en fonction des états internes des neurones et des poids synaptiques des connexions.

Processus : étape 3.2

- La sortie obtenue est comparée avec la sortie désirée pour l'exemple.
- On obtient alors une erreur, particulièrement importante en début d'apprentissage car la réponse du RNA est à ce stade purement aléatoire.

Processus : étape 3.3

Les poids synaptiques du RNA sont modifiés afin de réduire l'erreur calculée.

Processus : étape 3.4

On réitère cette opération un grand nombre de fois, et ce pour chaque exemple de la base, jusqu'à ce que le RNA converge vers une configuration qui lui permette de résoudre le problème à traiter.

4. Etape 4 : test

Afin de valider son fonctionnement, le RNA est testé sur des exemples n'ayant pas été préalablement appris.



Remarque

Généralement, faute d'une méthodologie affirmée, il est nécessaire de tester différentes modélisations et configurations d'apprentissage avant d'obtenir une résolution optimale du problème.

5. Etape 5 : utilisation



Méthode

- L'utilisateur introduit les informations sur le nouveau cas à résoudre.
- Par simple propagation des entrées à travers le réseau, le RNA fournit des sorties qui sont les solutions du problème.



Remarque

Ce processus est à la fois simple et rapide à réaliser.

C. Propriétés remarquables des RNA



Définition : Les propriétés remarquables d'un RNA sont :

- La capacité de généralisation
- La rapidité
- Une mémoire distribuée
- La résistance au bruit
- La résistance à la dégradation
- L'adaptabilité

1. La capacité de généralisation

- Le RNA sait résoudre plus qu'il n'a appris.
- Il généralise à partir des exemples qui lui ont été fournis à condition qu'ils soient nombreux, et est en mesure de donner des réponses sensées face à des cas de figures nouveaux.

2. La rapidité

- L'utilisation d'un réseau de neurones requiert une puissance informatique peu importante, l'essentiel ayant été fait lors de la phase d'apprentissage.
- La technique RNA est donc très rapide.

3. Une mémoire distribuée

- Une forme n'est pas mémorisée en un endroit précis (neurone ou connexion) d'un RNA.
- Comme dans le cerveau humain, la mémoire d'un fait est délocalisée en ce sens qu'elle est distribuée sur l'ensemble (ou tout au moins sur une partie) de la structure. Elle correspond à un état d'activation du réseau.



Remarque

La distribution de la mémoire confère aux RNA des propriétés importantes, comme :

- la résistance au bruit,
- ou la résistance à la dégradation.

4. La résistance au bruit

- Les RNA permettent de traiter des données dégradées, tant dans la phase d'apprentissage que dans la phase de présentation des problèmes.
- Des entrées entachées d'erreurs, imprécises ou incomplètes sont acceptables, à condition que la dégradation des données ne soit pas trop importante.



Remarque

Lorsqu'on lui injecte une forme quelque peu bruitée, un RNA reste à même de fournir la réponse attendue ou tout au moins une réponse correcte.

5. La résistance à la dégradation

L'expérience qui consiste à effectuer une petite lobotomie du réseau (un neurone par exemple) montre que les états mémorisés ne disparaissent pas pour autant ; tout juste peut-on constater une légère altération de la fonction de mémorisation.

6. L'adaptabilité

- Un RNA peut apprendre en cours d'utilisation.
- Il évolue et s'adapte au fur et à mesure des changements constatés sur des problèmes évolutifs.



Remarque

L'apprentissage ainsi réalisé est beaucoup plus simple que celui effectué avant utilisation et reste compatible avec une puissance informatique restreinte.

D. Domaine d'application

- Reconnaissance de formes
- Classification
- Identification
- Prédiction
- Filtrage
- Commande, régulation
- Optimisation combinatoire
- Diagnostic



Exemple : Reconnaissance de :

- Caractères (EasyReader),
- Visage (Mimetics),
- La parole,
- L'écriture (ordinateur sans clavier),
- Signature acoustique (Thompson),
- Objets (Silac : profilés d'aluminium)



Exemple : Diagnostic :

- Allumage voiture (Renault)
- Photocopieur (Canon)
- Circuits VLSI (IBM)



Exemple : Prédiction :

- De consommation d'eau (Lyonnaise)
- D'électricité (EDF)
- Trafic routier (Cofiroute)



VI



Le but du perceptron est d'associer des configurations (des formes) en entrée à des réponses.



- Méthode : Les cellules de la première couche répondent en oui/non.

- ### Remarque

- ## Example

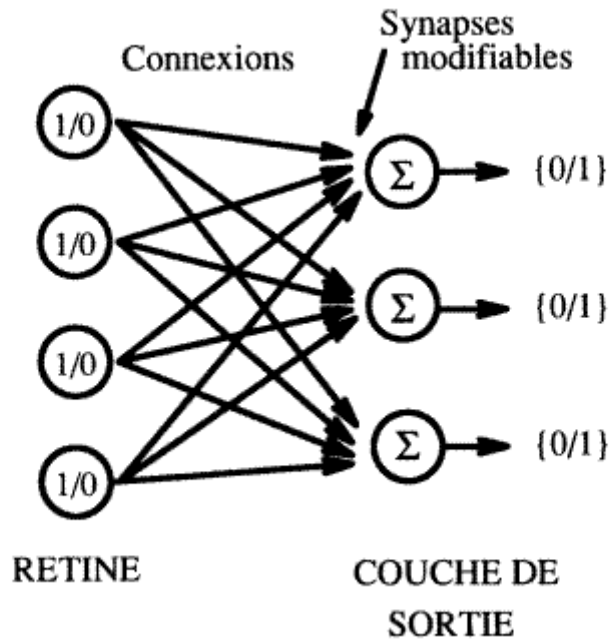


image8



Méthode

Les cellules de sortie évaluent l'intensité de la stimulation en provenance des cellules de la rétine en effectuant la somme des intensités des cellules actives.

La formule est : $A_j = \sum_i (X_i \cdot W_{ij})$

image9

A_j : activation totale de la jème cellule = potentiel post-synaptique

X_i : valeur (0 ou 1) de la ième cellule de la rétine

$W_{i,j}$: intensité de la connexion entre la ième cellule d'entrée et la jème cellule de sortie



Remarque

- L'activation en sortie est noté O_j .
- La fonction d'activation utilisée pour la sortie est appelée : fonction de Heaviside.
- Les cellules de sortie deviennent actives si leur degré d'activation (A_j) dépasse un seuil fixé (N_j).

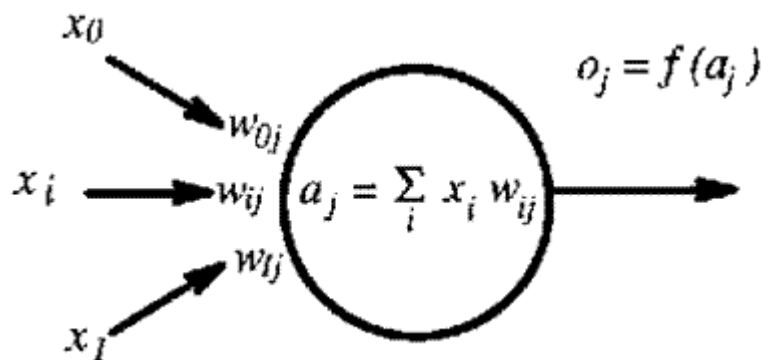


image10

$O_j = 0$ pour $A_j \leq N_j$

$O_j = 1$ pour $A_j > N_j$



Exemple : Perceptron qui donne le OU logique ?

????????????

Règle d'apprentissage

- Le perceptron doit trouver l'ensemble des valeurs à donner aux synapses pour que les configurations d'entrée se traduisent par des réponses voulues.
- Pour cela, on utilise la règle d'apprentissage de **Windrow-Hoff**.



Remarque

Pour apprendre, le perceptron doit savoir qu'il a commis une erreur, et doit connaître la réponse qu'il aurait dû donner.

=> De ce fait, on parle d'**apprentissage supervisé**.

La règle d'apprentissage est locale

Chaque cellule de sortie apprend sans avoir besoin de connaître la réponse des autres cellules.



Remarque

La cellule ne modifie l'intensité de ses synapses (apprend) que lorsqu'elle se trompe.



Méthode

- On va donner un ensemble de stimuli (intensité des synapses ou des connexions) au perceptron de façon arbitraire et on observe le résultat.
- Si le perceptron commet des erreurs, on lui permet de modifier l'intensité des connexions.
=> Les cellules du perceptron apprennent.
- Cette procédure est répétée jusqu'à ce que le perceptron soit capable de donner toutes les réponses correctes.



Définition : Règle d'apprentissage de windrow-Hoff

$$w_{i,j}^{(t+1)} = w_{i,j}^{(t)} + \eta(t_j - o_j) x_i = w_{i,j}^{(t)} + \Delta w_{i,j}$$

Avec:

X_i : valeur (0 ou 1) de la i ème cellule de la rétine

O_j : réponse de la j ème cellule de sortie

T_j : réponse théorique de la j ème cellule de sortie (0 ou 1)

$W_{i,j}$: intensité de la connexion entre la i ème cellule de la rétine et la j ème cellule de sortie au temps t (il faut savoir que les valeurs de $w(0)$ sont choisies arbitrairement)

η : constante positive dans le temps comprise entre 0 et 1. Le choix de η sera vu ultérieurement. Sa valeur influence le temps d'apprentissage. Dans certains cas, η varie en fonction du temps: on commence avec une valeur élevée et puis diminue à chaque itération.