

LES RESEAUX DE NEURONES APPLIQUES A LA RECHERCHE DOCUMENTAIRE

F. Z. BESSAI, A. NASRI, S. OULAD NAOUJ.

Laboratoire de Recherche et Développement en Informatique, (CERIST)

Rue des 3 frères Aissiou, Ben Aknoun, Alger, ALGERIE.

Fax : 213 (2) 91 21 26, Tél. : 213 (2) 91 18 21

E-mail: Bessai@tassili.cerist.dz

INTRODUCTION

Les modèles connexionnistes (neuronaux) sont des systèmes qui tentent de simuler les mécanismes de traitement de l'information ayant lieu dans le cerveau humain. Une des caractéristiques la plus importante des réseaux de neurones est leur capacité à apprendre: l'apprentissage va permettre au réseau de modifier sa structure interne (poids synaptiques) pour s'adapter à son environnement. En effet, nous n'avons pas besoin d'une formulation rigoureuse d'un problème donné pour le résoudre, tout ce dont nous avons besoin c'est une collection d'exemples représentative de la fonction désirée, ensuite le réseau s'adapte pour reproduire les sorties souhaitées quand un exemple d'entrée est présenté.

Une autre caractéristique des réseaux de neurones, contrairement aux méthodes algorithmiques traditionnelles, est leur capacité à traiter des données bruitées ou incomplètes. Un réseau de neurones, par sa structure composée de neurones ou cellules interconnectées entre elles par des liens synaptiques, est capable d'apprendre en modifiant ces liens. En plus il est doté de parallélisme massif dans son fonctionnement.

D'autres part, les Systèmes de Recherche Documentaire (SRD) développés jusqu'ici sont basés sur des modèles déclaratifs et des méthodes heuristiques. On peut dire d'une façon générale que se sont des systèmes statiques. La tendance actuelle dans l'informatique en général, et dans la recherche documentaire en particulier est le développement de systèmes dynamiques et évolutifs, capables d'apprendre et de modifier leurs comportements au fur et à mesure de leurs utilisations. Ceci est devenu possible grâce à l'intelligence artificielle et plus précisément à l'approche connexionniste qui explore la structure et le fonctionnement des réseaux neuronaux [11] [3] [9] [1] [10].

Dans cet article, nous commençons par une brève présentation des réseaux de neurones, leurs caractéristiques et leurs limites. Dans un deuxième temps, nous proposons un modèle de recherche documentaire utilisant la théorie des réseaux de neurones. Ce modèle présente un processus d'activation/propagation, des procédures de reformulation automatique des requêtes ainsi qu'un mécanisme d'apprentissage. Nous terminerons par une évaluation et une conclusion.

II. RESEAUX DE NEURONES

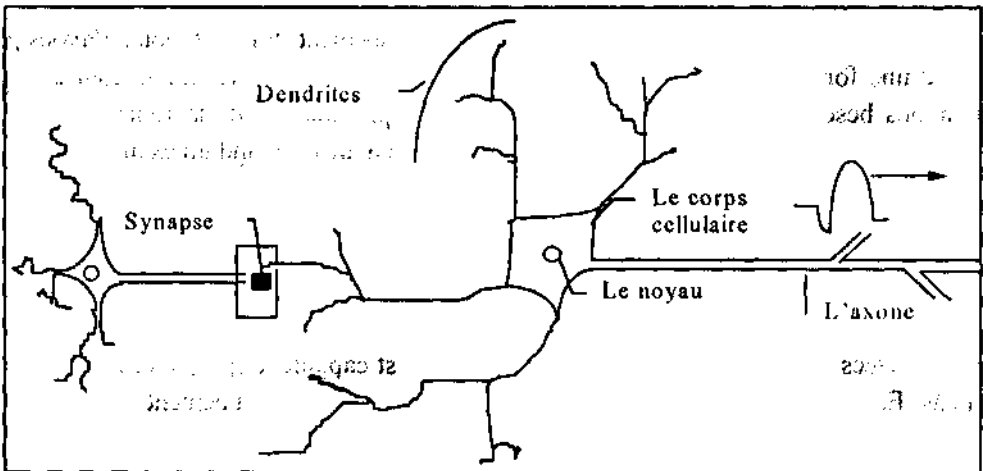
II.1. FONDEMENTS BIOLOGIQUES

Le cerveau est composé de milliards de cellules nerveuses appelées neurones et c'est l'activité électrochimique de ces cellules qui reflète l'accomplissement des tâches par le cerveau.

La description la plus classique des neurones est la suivante [11] [3] [9] [4] :

Chaque neurone possède une arborescence de fibres entrantes appelées dendrites, un élément de traitement appelé corps cellulaire et une arborescence de fibres sortantes appelée axone (voir Fig.II.1). La connexion de l'axone d'un neurone aux dendrites ou au corps cellulaire d'un autre neurone est appelé synapse.

Fig.II.1. Croquis simplifié d'un neurone biologique.



L'information circule sous forme de trains d'impulsions électriques. Le neurone, qui reçoit des informations depuis ses dendrites effectue approximativement l'intégration (la somme) de ces trains d'impulsions, si le potentiel somatique dépasse un certain seuil, il émet un signal bref que l'on appelle signal d'activation (voir Fig.II.2), ce dernier se propage sans amortissement le long de l'axone et des ramifications.

II.2. NEURONE FORMEL

La première modélisation d'un neurone a été présentée par Mac Culloch et Pitts en 1943. Ils ont proposé le modèle suivant : "Le neurone formel fait une somme pondérée des potentiels d'activation e_1, e_2, \dots, e_n qui lui parviennent, puis s'active suivant la valeur de cette sommation pondérée. Si cette somme dépasse un certain seuil, le neurone est activé et transmet une réponse, si le neurone n'est pas activé il ne transmet rien" [3] [13].

Ce modèle est schématisé par la figure suivante:

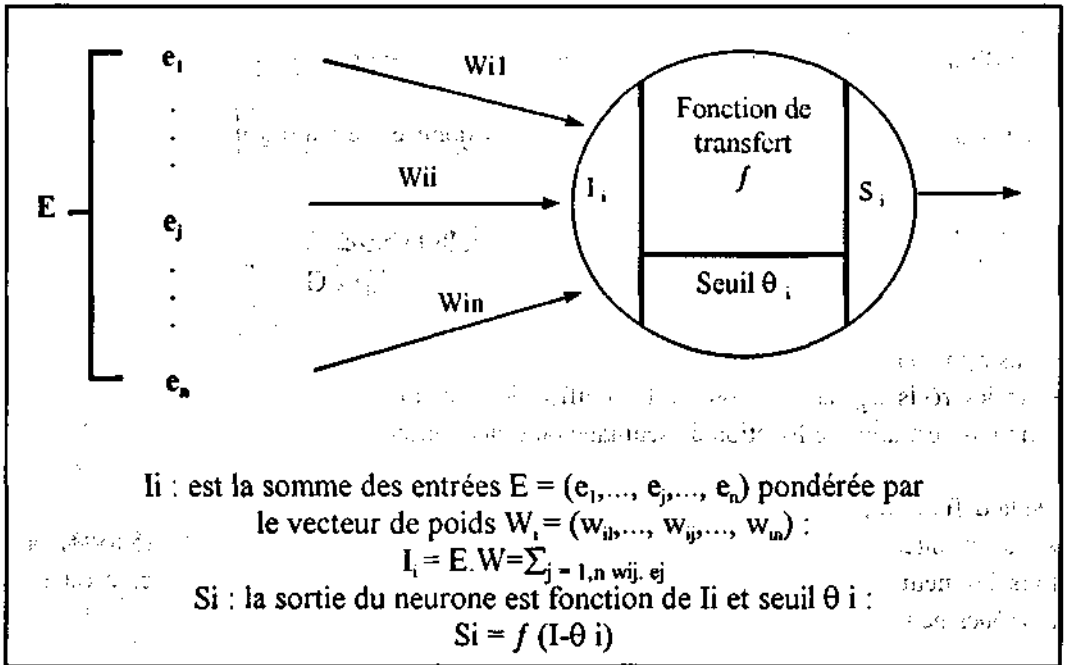


Fig.II.2. Le neurone formel.

La sortie du neurone dépend d'une fonction de transfert, dont les principales sont : fonction à seuil binaire (sign), fonction à seuil linéaire, fonction sigmoïde ($f(x) = \frac{e^x}{e^x + 1}$)

II.3. MODELISATION GENERALE D'UN RESEAU DE NEURONES

Un réseau de neurones est un ensemble de neurones formels (systèmes élémentaires) interconnectés et évoluant dans le temps par interactions réciproques. Pour décrire un tel réseau il nous faut donc décrire :

- le comportement de chaque neurone,
- l'interaction entre neurones.

Nous retiendrons la définition suivante :

II.3.1. DEFINITION STRUCTURELLE

Un réseau de neurones est défini à partir [WEN 12] :

- d'un ensemble I de N éléments appelés neurones, ou cellules, ou encore unités,
- d'un sous-ensemble Q de l'ensemble R des réels codant les différents états accessibles par un neurone,
- d'une fonction f de R dans Q ,
- d'un graphe (I,G) décrivant les connexions entre les neurones,
- d'un réel w_{ij} , pour chaque couple (i,j) de G ,
- d'un ensemble d'entrées E ,
- et d'une fonction i de E dans R , pour chaque élément i de I .

Un réseau de neurones construit à partir de ces données, est alors un système dont la réponse à une entrée e de E est un vecteur S de R^n vérifiant :

$$s_i = f \left(\sum_j w_{ij} s_j + \theta_i(e) \right) \text{ pour } \Rightarrow \Leftarrow \text{ toute } \Rightarrow \Leftarrow \text{ cellule } \Rightarrow \Leftarrow i, \text{ j tel que } (i,j) \in G$$

Nous appelons, respectivement, synapses et coefficients synaptiques les éléments (i,j) de G et les réels w_{ij} , la fonction i est qualifiée de seuil lorsqu'elle est constante sur E , la fonction f est appelée fonction de seuillage ou d'activation.

De la définition précédente nous tirons les remarques suivantes :

- les architectures utilisées sont très variées, peuvent aller d'une connectivité totale où tous les neurones sont connectés les uns aux autres à une connectivité locale où les neurones ne sont connectés qu'à leurs proches voisins.
- il est courant d'utiliser des réseaux à structures régulières pour faciliter leur utilisation [3].
- le coefficient synaptique (w_{ij}) appelé aussi poids, attribué à la connexion entre deux neurones (i,j) de G , a pour rôle de rendre compte de l'influence plus ou moins grande d'un neurone sur ceux auxquels est reliée sa sortie.

II 3.2. DYNAMIQUE DES ETATS

La dynamique des états s'intéresse à l'évolution des états des différents neurones d'un réseau. Elle cherche l'existence d'états stables ou de cycles stables, que ce soit pour des cellules particulières, pour des groupes de cellules ou pour le réseau tout entier [3].

Elle dépend de:

- la fonction d'activation des neurones,
- la structure et poids des connexions.

II.3.3. DYNAMIQUE DES CONNEXIONS

Le poids w_{ij} de la connexion entre les neurones (i,j) pondère le signal transmis, est peut être :

- excitateur ($w_{ij} > 0$)
- inhibiteur ($w_{ij} < 0$)

La dynamique des connexions ne s'intéresse qu'aux réseaux dont les poids des connexions sont évolutifs [3] [9].

II.3.4. FONCTIONNEMENT

Une fois l'architecture et la dynamique du réseau sont choisies, le réseau va subir à son entrée les motifs à apprendre (phase d'apprentissage), l'algorithme d'apprentissage détermine la façon d'ajuster les poids du réseau pour obtenir la sortie désirée pour un motif donné. La phase suivante est appelée phase d'utilisation ou de rappel, elle consiste à présenter au réseau des motifs autres que ceux qui ont contribué à son apprentissage (des motifs généralement bruités ou incomplets), pendant cette phase le réseau va réagir selon les connaissances acquises durant la phase d'apprentissage.

II.4 APPRENTISSAGE

Une des caractéristiques les plus intéressantes des réseaux de neurones est leur capacité à apprendre, l'apprentissage va permettre au réseau de modifier sa structure interne (poids synaptiques) pour s'adapter à son environnement [11] [3] [9].

II.4.1. PRINCIPE

Dans le cadre des réseaux de neurones un réseau est défini par son graphe de connexions et la fonction d'activation de chaque neurone.

A chaque choix de coefficients synaptiques (poids de connexions) correspond alors un système, et c'est dans l'ensemble de tous ces systèmes que l'on se propose de trouver celui résolvant au mieux le problème [12].

Pour pouvoir évaluer un système particulier, nous effectuons pour cela une suite d'expériences permettant d'observer le comportement du réseau. Une expérience consiste à présenter un exemple d'entrée au système dont la réponse est recueillie à la sortie du système.

Le réseau est ainsi évalué par la valeur prise par une fonction d'erreur. Le problème d'apprentissage consiste alors à trouver un réseau minimisant cette fonction d'erreur [12].

II.4.2. MODES D'APPRENTISSAGE

Nous distinguons trois modes d'apprentissage : l'apprentissage supervisé, non supervisé et renforcé.

II.4.2.1. APPRENTISSAGE SUPERVISE

Dans ce mode, un professeur qui connaît parfaitement la sortie désirée ou correcte guide le réseau en lui apprenant à chaque étape le bon résultat. Donc l'apprentissage ici consiste à comparer le résultat obtenu avec le résultat désiré, puis à ajuster les poids des connexions pour minimiser la différence entre les deux [11].

II.4.2.2. APPRENTISSAGE NON SUPERVISE

Dans l'apprentissage non supervisé, le réseau modifie ses paramètres en tenant compte seulement des informations locales. Ces méthodes n'ont pas besoins de sorties désirées préétablies. Les réseaux utilisant cette technique sont appelés réseaux à dynamique autonome et sont considérés comme des détecteurs de régularité, car le réseau apprend en détectant les régularités dans la structure des motifs d'entrée et produit la sortie la plus satisfaisante [3].

II.4.2.3. APPRENTISSAGE RENFORCE

Il est utilisé quand une information en retour sur la qualité de la performance est fournie, mais que la conduite souhaitée du réseau n'est pas complètement spécifiée par un professeur. Donc l'apprentissage est moins dirigé que l'apprentissage supervisé. Contrairement à l'apprentissage non supervisé où aucun signal de retour n'est donné, le réseau à apprentissage renforcé peut utiliser le signal de renforcement pour trouver les poids les plus désirables quand c'est nécessaire [11].

II.4.3. REGLES D'APPRENTISSAGE

La méthode d'ajustement des poids du réseau pendant l'apprentissage peut être choisie dans une gamme variée d'algorithmes, dont voici les plus connus :

- la règle de Hebb (ou règle de corrélation) [5] [11]
- la règle de Widrow-Hoff [4] [11] [7]
- la règle delta
- l'algorithme de rétropropagation de l'erreur (ou la règle delta généralisée) [11]
- apprentissage compétitif [7] [4]

II.4.4. PROPRIETES ET LIMITES DES RESEAUX DE NEURONES

II.4.5. LES PROPRIETES FASCINANTES DES RESEAUX DE NEURONES

L'intérêt porté aujourd'hui aux réseaux de neurones tient sa justification dans les quelques propriétés suivantes [10]:

II.4.5.1. LA CAPACITE D'ADAPTATION

La capacité d'adaptation se manifeste dans les réseaux de neurones par la capacité d'apprentissage qui permet au réseau de tenir compte de nouvelles contraintes ou de nouvelles données. Cette capacité présente un intérêt déterminant pour les problèmes volutifs.

II.4.5.2. LA CAPACITE DE GENERALISATION

La capacité de généralisation se traduit par la capacité d'un système à apprendre et à retrouver, à partir d'un ensemble d'exemples, des règles sous-jacentes ou à mimer des comportements qui permettent de résoudre un problème.

II.4.5.3. LE PARALLELISME

Cette notion se situe à la base de l'architecture des réseaux de neurones considérés comme un ensemble d'entités élémentaires qui travaillent simultanément. Le parallélisme permet une rapidité de calcul supérieure mais exige de penser et de poser différemment les problèmes à résoudre.

II.4.6. LIMITES D'UTILISATION DES RESEAUX DE NEURONES

Ces limites [3] [11] sont tout d'abord d'ordre technique et sont dues aux difficultés que nous rencontrons pour utiliser le parallélisme inhérent aux réseaux de neurones. Ainsi, la plupart des réseaux sont simulés sur des machines séquentielles, ce qui entraîne des temps de calculs importants.

De plus, l'un des principaux reproches fait aux réseaux de neurones tient de leur incapacité à expliquer les résultats qu'ils fournissent. Les réseaux se présentent comme des boîtes noires dont les règles de fonctionnement sont inconnues. La qualité de leurs performances ne peut être mesurée que par des méthodes statistiques, ce qui amène une certaine méfiance de la part des utilisateurs potentiels.

II.5. CONCLUSION SUR LES RESEAUX DE NEURONES

Un réseau de neurones est un ensemble de cellules interconnectées par des liens ajustables. Cette propriété lui permet de trouver une représentation interne d'un problème donné, le réseau grâce à cette représentation, est capable de reproduire la sortie appropriée pour une entrée qui lui est présentée.

Les propriétés des réseaux de neurones ont permis à ces derniers de trouver une large utilisation dans divers domaines d'application tels que, la reconnaissance de formes, le traitement de signal, la robotique, le traitement de l'information, l'optimisation combinatoire, etc...

Rappelons que pour valider l'adaptation d'un problème donné à une solution neuronale, il est nécessaire :

1. que le problème présente certaines caractéristiques parmi les suivantes :
 - a) Les règles qui permettraient de résoudre le problème sont inconnues ou très difficiles à expliquer ou à formaliser. Cependant, nous disposons d'un ensemble d'exemples qui correspondent à des entrées du problème et aux solutions qui leur sont données par des experts ,
 - b) Le problème fait intervenir des données bruitées ,
 - c) Le problème peut évoluer ,
 - d) Le problème nécessite une grande rapidité de traitement ,
 - e) Il n'existe pas de solutions technologiques courantes.
2. de pouvoir disposer d'un nombre suffisant de données pour assurer à la fois l'apprentissage et la vérification des performances du réseau,
3. d'étudier la nature des données qui peuvent nécessiter un prétraitement.

Toutefois, l'expertise, à la fois dans le domaine d'application et dans les réseaux de neurones, est nécessaire pour la réussite de la modélisation connexionniste et de sa mise en oeuvre. Ainsi, l'application des réseaux de neurones au domaine de la recherche documentaire se justifie par les points suivants :

1. le parallélisme devient une nécessité pour améliorer le temps de réponse dans les bases de données volumineuses,
2. grâce aux mécanismes d'activation/propagation, le système offre des moyens pour reformuler les requêtes des utilisateurs et de répondre aux requêtes incomplètes ou mal formulées grâce à la capacité de généralisation ,
3. avec la capacité d'adaptation , le système modifie sa structure interne afin de satisfaire les désirs de chaque utilisateur.

Dans ce qui suit, nous allons proposer un modèle connexionniste pour la recherche documentaire en terme d'architecture et de dynamique de fonctionnement du réseau.

Le langage utilisé pour la modélisation est le langage de programmation PROLOG. Ce langage est particulièrement adapté à la modélisation de la recherche documentaire.

III. UN MODELE CONNEXIONNISTE POUR LA RECHERCHE DOCUMENTAIRE

III.1. MODELISATION DE LA BASE D'INFORMATIONS

Le modèle que nous proposons pour représenter la base de documents est formé de trois couches [8] : - la couche requête, notée Q_k ,
- la couche termes, notée T ,
- la couche documents, notée D .

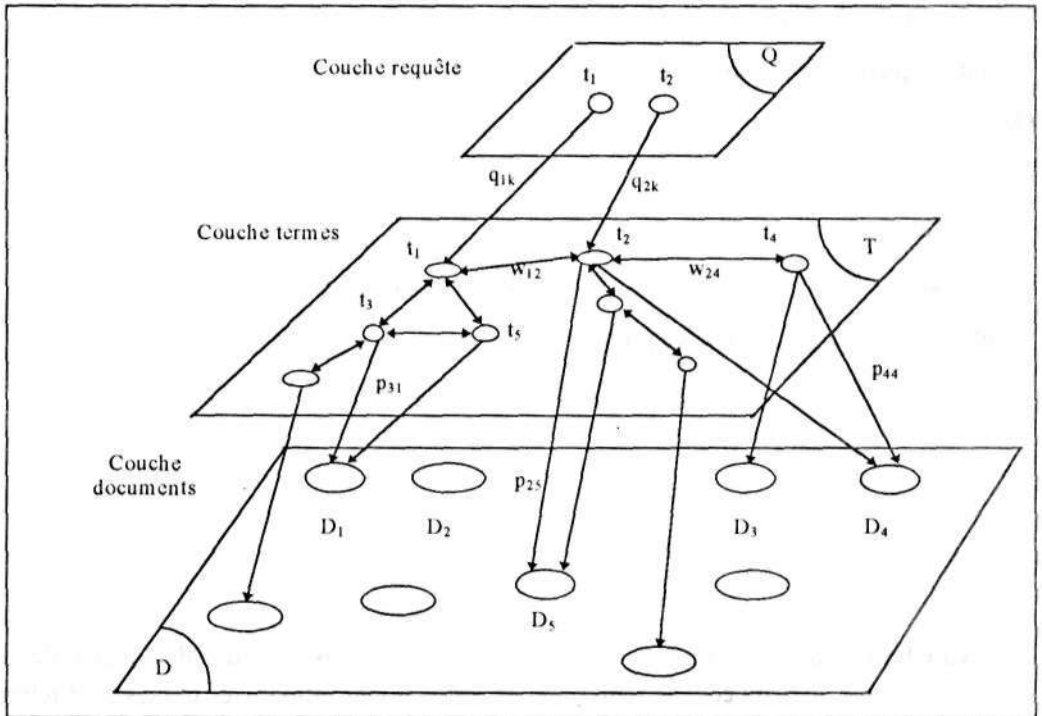


Fig. III.1: le réseau de neurones modélisant la base de documents

1. la couche requête, Q_k , cette couche représente l'entrée du réseau, elle est créée dynamiquement à chaque interrogation. L'existence du neurone i sur cette couche implique la présence du terme t_i dans la requête. Il existe un lien synaptique q_{ik} reliant chaque terme de la requête à un terme de la couche T reflétant le poids de ce terme dans la requête Q_k . Ce poids est donné par la formule :

$$q_{ik} = \sum_{j=1}^M \frac{p_{ij}}{M} \quad (1)$$

Où :

q_{ik} : Le poids absolu du terme t_i

p_{ij} : Le poids relatif du terme t_i au document D_j ,

M : Le nombre de documents contenant t_i .

Chaque requête peut être représentée comme suit :

$$Q_k = (a_{1k}, \dots, a_{Tk})$$

Où :

Q_k : est la requête numéro k ,

a_{ik} : est un booléen indiquant la présence du terme t_i dans la requête Q_k .

T : le nombre de termes dans le système.

$$a_{ik} = \begin{cases} 1 & \text{si } t_i \in Q_k \\ 0 & \text{sinon} \end{cases} \quad (2)$$

2. la couche termes, T , représente l'ensemble des termes extraits des documents de la base. Une cellule de cette couche représente un terme d'indexation. Ces termes sont reliés entre eux par des liens w_{ij} représentant la concurrence des deux termes t_i et t_j . Ces liens sont donnés par la formule :

$$w_{ij} = \frac{\text{card}(E_{t_i} \cap E_{t_j})}{\text{card}(E_{t_i} \cup E_{t_j})} \quad (3)$$

Où :

E_{t_k} : L'ensemble de documents contenant le terme t_k ,

card : Le cardinal d'un ensemble

Nous pouvons donc déduire de la formule (3) que : $\forall_{i,j} \in [1, T]$:

$$0 \leq w_{ij} \leq 1$$

Où T : est le nombre de termes dans le système.

Chaque terme de cette couche est relié à la couche documents par des liens p_{ij} valués, tel que m appartient à l'ensemble des descripteurs des documents où ce terme apparaît. A l'état initial ces liens sont donnés par la formule :

$$P_{ij} = \frac{\sqrt{N}}{N - n_i} \quad (4)$$

Où :

p_{ij} : Poids du terme t_i dans la caractérisation du contenu du document D_j ,

n_i : Le nombre de descripteurs de documents contenant le terme t_i ,

N : Le nombre total de documents dans la base.

3. la couche documents, D, c'est la sortie du réseau, elle représente les documents de la base.

Ces différents liens inter et intra-couches sont initialisés à la phase d'indexation puis ajustés à leurs valeurs optimales lors de l'apprentissage.

Le fonctionnement du réseau est divisé en deux phases :

- la phase recherche,
- la phase apprentissage.

III.2. Recherche associative d'informations

Dans cette phase le réseau reçoit à son entrée (couche requête) un vecteur de termes activant ainsi les termes de la requête sur la couche T , ensuite ces termes activés sur T vont propager leurs activations à leurs proches voisins. En fin, les termes activés directement à partir de la couche requête et ceux activés par propagation, vont envoyer leurs signaux d'activation vers la couche D à travers les liens p_{ij} . Les documents sur la couche D recevant des signaux suffisants pour être activés, s'activent, et constituent la réponse à la requête de l'utilisateur, comme l'illustre le schéma suivant (figure III.2).

La fonction d'activation que nous avons choisi, pour toutes les couches, est la fonction sigmoïde [8] [2] [12] [4] :

$$f(x) = \frac{e^x - 1}{e^x + 1} \quad (5)$$

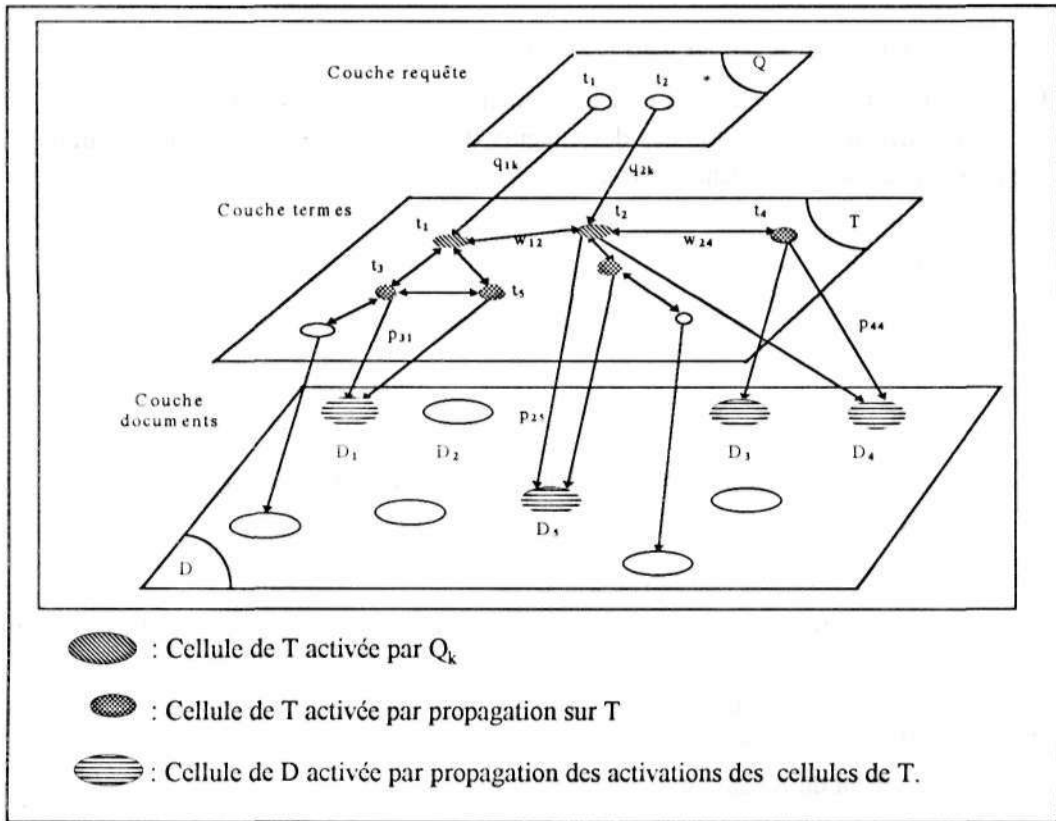


Fig. III.2: Activation/propagation induite par une requête

Nous distinguons deux types de propagations :

- Propagation directe (sans extension de la requête),
- Propagation aux autres unités cachées de la couche des termes, ce qui permet l'extension de la requête.

Ces deux types de propagation permettent d'avoir deux modes de recherche :

- recherche sans extension de la requête,
- recherche avec extension de la requête.

Recherche sans extension de la requête

Le processus de recherche est effectué comme suit:

1. le vecteur de la requête Q_k représentant le signal d'entrée sera propager à la couche termes, où chaque neurone de cette couche calcule son entrée selon la formule suivante :

$$E^T(t_i) = \sum_{j=1}^T a_{jk} \cdot q_j \quad (6)$$

Vu la définition des booléens aik tous les termes de la somme dans la formule (6) vont s'annuler sauf le terme correspondant à $j=i$ d'où :

$$E^T(t_i) = q_i \quad (7)$$

Chacun de ces termes sera ensuite activé selon l'équation :

$$S^T(t_i) = f(E^T(t_i)) \quad (8)$$

Où : f est la fonction sigmoïde.

2. Chaque nœud activé de la couche T propage sa sortie à travers les liens p_{ij} vers la couche documents. Les neurones de cette couche vont à leur tour évaluer leurs entrées pondérées selon la formule :

$$E^D(D_j) = \sum_{i=1}^T S^T(t_i) \cdot P_{ij} \quad (9)$$

Puis s'activeront selon la formule :

$$S^D(D_j) = f(S^D(D_j)) \quad (10)$$

Les documents activés sont triés selon leurs valeurs d'activation décroissante.

Sur la couche documents nous utilisons un seuil, à partir duquel un nœud document sera activé de la façon suivante :

$$D_j = \begin{cases} \text{actif si } SD(D_j) \leq \theta \\ \text{inactif sinon} \end{cases} \quad (11)$$

Où : θ : est le seuil d'activation.

III.2.2. Recherche avec extension de la requête

L'extension de la requête est faite grâce aux liens de concurrence entre les termes. Dans ce mode de recherche, un nœud terme évalue son activation, et l'envoie à ses proches voisins à travers les liens de concurrence, ce qui implique la reformulation automatique de la requête sur la couche T.

Les nœuds termes activés par propagation calculent leurs entrées par :

$$E^T(t_i) = \sum_{j=1}^T S^T(t_j) \cdot w_{ij} \quad (12)$$

et seront activés selon la formule :

$$S^T(t_j) = f(E^T(t_j)) \quad (13)$$

Chaque document calcule son entrée selon la formule (9) et s'active selon la formule (10). Comme précédemment, les documents activés représentent la réponse du système au besoin de l'utilisateur. Ces documents sont ordonnés selon leurs valeurs d'activation décroissante. L'utilisateur sélectionne les documents qu'il juge pertinents, et peut, ensuite, déclencher le processus d'apprentissage.

III.3. Apprentissage

L'apprentissage constitue une nouvelle ouverture avec l'utilisateur. En effet, en tenant compte des jugements des utilisateurs, l'apprentissage modifie les liens du réseau pour répondre au mieux aux besoins de ces utilisateurs. L'idée, d'une façon générale, est de renforcer les liens qui collaborent positivement pour approcher la pertinence et de diminuer les liens qui participent négativement à cette pertinence.

La phase d'apprentissage ne peut être déclenchée qu'après avoir induit le système dans une phase de recherche préalable. Durant cette phase, l'utilisateur va désigner les documents pertinents, et à partir de cette information, le système modifie les poids des liens et/ou crée de nouveaux liens entre les noeuds du réseau afin d'améliorer la réponse déjà donnée par le système. L'utilisateur peut alors refaire la phase de recherche ou d'apprentissage en donnant la requête initiale et détecte ainsi l'effet de son jugement à travers la réponse du système.

Avant d'aborder de manière détaillée l'apprentissage sur les différentes couches du réseau une règle générale de modification des poids sera proposée.

- Règle générale de modification des poids :

Les règles d'apprentissage que nous proposons sont inspirées de l'algorithme d'auto-organisation de Kohonen [6] et de la règle de Hebb [5] [4] [7].

L'algorithme de Kohonen est utilisé pour réaliser les opérations de classification et de regroupement, tandis que la règle de Hebb est utilisée pour simuler le fonctionnement du neurone biologique.

La règle générale proposée est la suivante :

$$\text{poids}_{ij}(t+1) = \text{poids}_{ij}(t) + (x_{ij} - \text{poids}_{ij}(t)) \quad (14)$$

Où :

$$X_{ij} = \begin{cases} 1 & \text{si } i \text{ et } j \text{ sont deux neurones représentant deux concepts pertinents (a)} \\ 0 & \text{sinon (b)} \end{cases} \quad (15)$$

La formule (14) se caractérise par les trois points suivants :

- permet au réseau de former des amas de neurones selon l'algorithme d'auto-organisation de Kohonen,
- s'accorde avec la règle de Hebb, du fait qu'elle augmente le lien entre deux concepts pertinents et actifs en même temps et diminue ce lien dans le cas contraire et,
- assure la plasticité/stabilité des poids synaptiques dans l'intervalle [0,1], qui peuvent dans le cas de l'application de la règle de Hebb pure diverger en dehors de cette intervalle. N'oublions pas que les systèmes biologiques ont leurs propres mécanismes de conserver leurs énergies. Cette énergie est donnée dans notre cas, par exemple sur la couche T, par la formule suivante :

$$E = -1/2 \sum_{j \neq i} w_{ij} S^i(t_j) \cdot S^j(t_j) \quad (16)$$

Le réseau tend à trouver des attracteurs (des minima) de cette énergie où se situe le maximum des réponses pertinentes aux besoins de l'utilisateur.

Le modèle que nous avons proposé a conduit à la réalisation d'un système de recherche documentaire, appelé SYCORD (Système Connexionniste de Recherche Documentaire), dont l'évaluation se présente dans la section suivante.

IV. EVALUATION

Pour l'évaluation du système SYCORD nous avons utilisé un échantillon de la base INSPEC (Information Services for Physics and Engineering Communities) dont les caractéristiques sont les suivantes:

Nombre de documents	3123
Nombre de termes	3095
Nombre de liens w_{ij}	20248
Nombre de liens p_{ij}	12461

Tableau 1. L'état initial du système (avant apprentissage)

Les expérimentations qui ont été réalisées concernent un utilisateur donné pendant l'évaluation du système dans l'ordre suivant :

1. recherche sans reformulation,
2. recherche avec reformulation,
3. recherche avec reformulation de la requête et apprentissage.

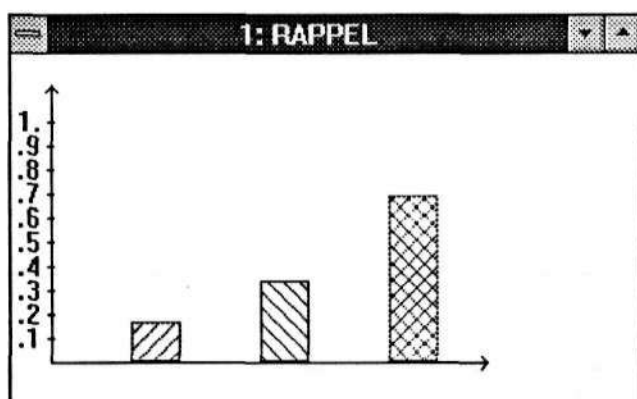
Après chaque processus d'évaluation, nous constituons les graphes 1, 2 et 3 donnés ci-dessous pour comparer les résultats obtenus dans chacun des trois cas précédents.

Le processus d'apprentissage a permis d'améliorer les résultats de recherche. En effet, nous avons remarqué qu'après apprentissage :

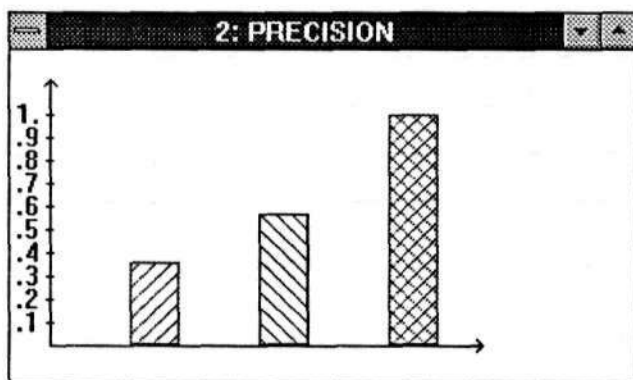
1. le taux de rappel est amélioré, car la valeur d'activation d'un document est proportionnelle à la valeur des liens de cooccurrence entre les termes qui lui sont reliés, donc le fait d'avoir renforcé les liens entre les termes de la requête et les termes reliés à des documents pertinents a permis au système de sélectionner des documents pertinents en plus. Avant l'apprentissage ces documents n'ont pas pu être sélectionnés car leurs activations étaient inférieures au seuil (voir graphe 1)

2. le taux de précision a également augmenté, en effet, les documents non pertinents sont au fur et à mesure éliminés par le processus d'apprentissage (voir graphe 2)

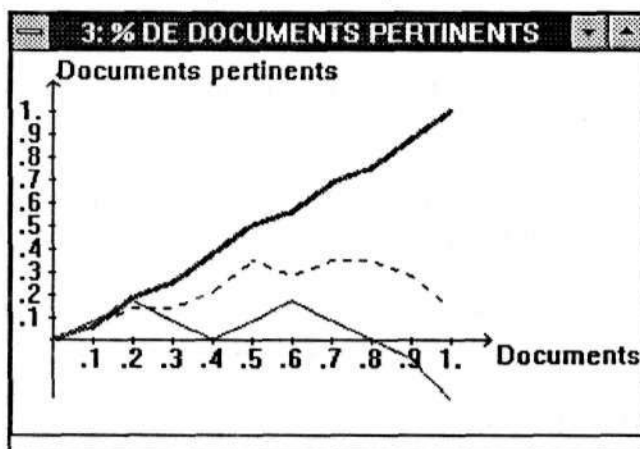
3. la distribution et le nombre des documents pertinents par rapport aux documents non pertinents sont améliorés (les documents jugés pertinents se trouvent au début de la liste de documents sélectionnés) (voir graphe 3)



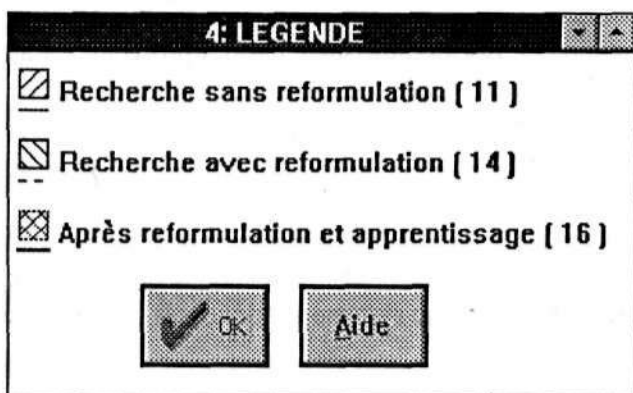
Graphe 1. Taux de rappel



Graphe 2. Taux de précision



Graphe 3. Distribution des documents



Les nombres entre parenthèses figurant dans cette légende donnent le nombre de documents sélectionnés dans chaque phase de recherche.

Nombre de documents	3123
Nombre de termes	3095
Nombre de liens w_{ij}	20218
Nombre de liens p_{ij}	12461

Tableau 2. L'état du système après apprentissage

Remarque: Il y a variation du nombre de liens w_{ij} due au processus d'apprentissage.

Les résultats de cette évaluation montrent comment le système apprend et réagit aux besoins d'un utilisateur.

CONCLUSION :

A l'issue de ce travail, nous avons montré la faisabilité de l'approche connexionniste pour les problèmes de recherche documentaire.

Les caractéristiques principales de notre modèle se situent autant dans la souplesse de son architecture que dans ses dynamiques de fonctionnement et d'évolution. Ainsi, concernant l'architecture, la représentation de la base en couches facilite la mise en oeuvre d'une part, des formules d'activation/propagation propres à chaque type de liens et d'autre part, les règles d'apprentissage permettant l'évolution de la base.

Pour ce qui est de la dynamique de fonctionnement, elle a pour but de rechercher des informations. Cette recherche, qualifiée de recherche "parallèle" ou "associative", réalisée grâce à un processus d'activation/propagation utilisant les différents liens du réseau.

La dynamique d'évolution du réseau est assurée par des procédures d'apprentissage. Ces procédures consistent d'une façon générale à modifier les poids des différentes connexions, et plus précisément à établir et à affecter des poids sémantiques reflétant l'importance effective de chaque terme dans la caractérisation du contenu de chaque document.

En fin, nous pouvons dire que le système SYCORD ainsi réalisé nous a permis de faire une validation expérimentale en plus de la validation théorique que nous avons justifiée dans cet article. Il reste bien sûr à élargir ce système sur des bases plus importantes.

Références Bibliographiques

- [1] **I.Aleksander, H.Morton**
An introduction to neural computing
CHAPMAN A.HALL 1990
- [2] **M.Boughanem, C.Soùle-dupuy**
Query expansion and neural network
IPMM, 1994
- [3] **E.Davalo, P.Naïm**
Des réseaux de neurones
Eyrolles 1990
- [4] **J.A.Freeman, D.M.Skapura**
Neural Networks
Algorithms, Application and Programming techniques
Addison Wesley 1992.
- [5] **D.O.Hebb**
The organisation of behavior
J.Wiley & Sons NY, 1949
- [6] **T.Kohonen**
Self-organizing maps: optimization approaches
ICANN'91 Finland
- [7] **P.Martin**
Réseaux de neurones artificiels
Application à la reconnaissance optique de partitions musicales
Thèse de doctorat université de J.F.Grenoble I 1992.
- [8] **A. NASRI, S. OULAD NAOUI**
Conception et réalisation d'un système de recherche documentaire utilisant
l'approche connexionniste (SYCORD)
Mémoire d'ingénieur INI (Institut National de formation en Informatique)
1996

[9] M. Mc Cord Nelson, W.T. Illingworth

A practical guide to neural nets

Addison Wesley Texas instruments 1990.

[10] G.N.Reddy

Artificial neural networks

Theory, Computer simulation programs and Applications

Dept. of Electrical Engineering Lamar university

Beaumont, Texas, USA Nov. 1995

[11] H.Samelides, P.Bouret, J.Reggia

Réseaux neuronaux

une approche connexionniste de l'intelligence artificielle.

édition TEKNEA 1991.

[12] D.Wenzek

Construction de réseaux de neurones

Thèse de doctorat INPG 1993.

[13] J. A. Farrel, A. N. Michel

Associative memory via artificial neural networks

IEEE control system magazine 1990.