

RESEAUX DE NEURONES ET PSYCHOMETRIE

Etude prospective des applications possibles des réseaux de neurones formels dans le traitement des données psychométriques.

Editions du Centre de Psychologie Appliquée

Benoît Virole

Juin 2001

SOMMAIRE

1. Introduction aux réseaux de neurones	3
1.1 Présentation.....	3
1.2 Définitions	3
1.3 Importance de la fonction de seuil	5
1.4 Structure général d'un réseau de neurones.....	6
1.5 Les mécanismes d'apprentissage	7
1.6 Architecture des principaux réseaux de neurones.....	8
2. Applications en psychométrie	13
2.1 Avantages et inconvénients des réseaux de neurones	13
2.2 Indications des réseaux de neurones en psychométrie	14
3. Etude sur la prédiction d'une catégorie externe	15
3.1 L'échelle BDI-II.....	15
3.2 Méthodologie générale	16
3.3 Résultats	17
4. Etude sur la classification automatique	22
4.1 Utilisation de la carte topologique de Kohonen.....	22
4.2 Procédure standard en classification automatique	26
5. Etude sur la prédiction de données manquantes.....	27
5.1 Le problème des données manquantes.....	27
5.2 Procédure d'imputation par réseau de neurones	30
6. Etudes en cours utilisant des réseaux de neurones	32
6.1 Etude longitudinale sur les implants cochléaires chez l'enfant	32
6.3 Evaluation des risques psychopathologiques chez les enfants de patients traumatisés crâniens.....	33
7. Conclusions	34
Bibliographie	35

1. Introduction aux réseaux de neurones

1.1 Présentation

Disponibles aujourd'hui au sein de la plupart des grands logiciels de statistique, les modules de réseaux de neurones permettent d'effectuer des traitements statistiques originaux. L'objectif de ce document est de présenter l'intérêt de ces réseaux de neurones dans le traitement de données psychométriques. Sans prétendre remplacer les analyses conventionnelles, ils peuvent rendre des services importants tant sur le plan pratique que sur le plan théorique. Ils ouvrent également des perspectives dans la prédiction et peuvent ainsi fournir aux tests psychologiques un complément fonctionnel intéressant.

Ce document présente la théorie des réseaux de neurones formels et les résultats d'études destinées à tester l'efficacité de leur utilisation dans le traitement de données psychométriques. L'idée de départ est qu'il est possible d'utiliser la propriété de convergence des réseaux de neurones formels. Selon cette propriété fondamentale, un réseau peut apprendre à reconnaître le noyau central d'un nuage de données et évaluer la distance de chaque point vis-à-vis de ce noyau que l'on désigne sous le nom d'attracteur du réseau. La première étude teste la convergence entre les données d'un étalonnage mises en variables d'entrée et une variable externe concernant une catégorisation clinique. On apprend au réseau à identifier une catégorie clinique sur un nombre de sujets appartenant à l'étalonnage. Puis on vérifie sur une autre partie des sujets (sélectionnés au hasard) la capacité du réseau à identifier la catégorie. Le réseau est alors l'image des liens entre les deux ensembles. Une autre étude (BDI-II) vise à tester la capacité des réseaux de neurones à extraire de l'information à partir de données d'entrée au travers d'une classification automatique. Une troisième étude traite de la capacité d'imputation des données manquantes. En liant les données de test à une variable de test qu'on positionne en variable externe, on teste la capacité du réseau à restituer des données manquantes (RCMAS). Enfin, on présentera brièvement deux études psychométriques en cours. Une utilise les réseaux de neurones pour la prédiction de réussite des implants cochléaires chez les enfants sourds à partir des données du PSA et une autre étude traite des relations entre les données du RCMAS et des données cliniques chez les enfants de patients traumatisés crâniens..

1.2 Définitions

Par analogie avec les neurones biologiques, on appelle réseaux de neurones des ensembles de calculateurs numériques qui agissent comme des unités élémentaires. Ils sont reliés entre eux par un ensemble d'interconnexions pondérées qui transmettent des informations numériques d'un « neurone formel » à l'autre. L'apprentissage du réseau de neurones consiste alors à modifier répétitivement les valeurs de pondération des interconnexions jusqu'à ce que la performance du réseau atteigne le niveau désiré. Un neurone formel reçoit en entrée des informations sous la forme de valeurs numériques, d'en faire un traitement, et selon ses résultats, de générer une information de sortie vers un autre neurone formel (cf. figure 1.). Chaque neurone formel possède des caractéristiques

propres, en particulier un seuil de déclenchement, assimilable à un poids synaptique dont le dépassement implique la décharge du neurone, c'est-à-dire la transmission d'une information de sortie.

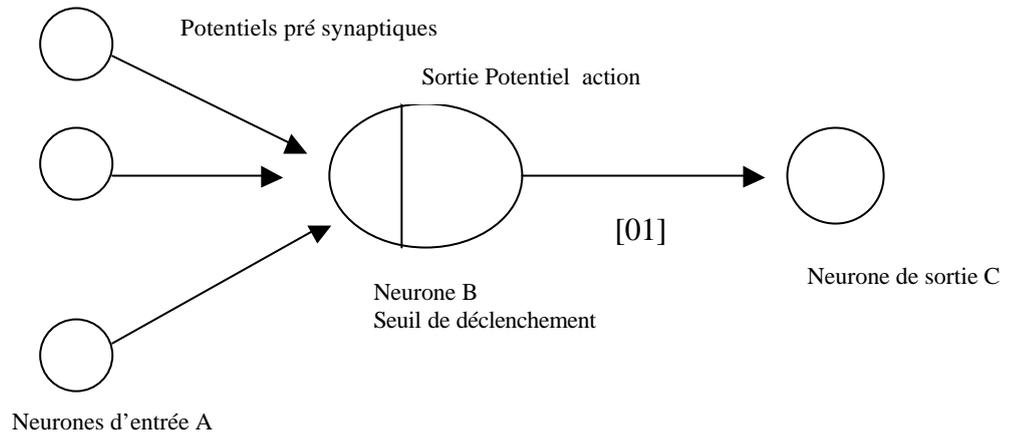


Figure 1. Schéma général d'un neurone formel

Les neurones d'entrée envoient des valeurs numériques au neurone de réception qui en effectue une somme numérique (éventuellement pondérée). Si cette valeur dépasse le seuil (fixé par défaut mais modifiable ensuite par apprentissage) alors le neurone B envoie une information (1) au neurone C. Dans le cas où la sommation est inférieure au seuil, le neurone B envoie une information (0) au neurone C. Ce processus se reproduit dans toutes les interconnexions entre les neurones du réseau.

Sur la figure 1, le cercle large représente le corps cellulaire (soma). Le soma est relié par un certain nombre d'entrées. Chaque entrée est une combinaison d'une dendrite et d'une synapse. Les canaux d'entrée dans le soma sont activés par des signaux qu'ils reçoivent des neurones auxquels ils sont connectés. Le potentiel post-synaptique, déclenché sur le neurone formel est le résultat d'une somme pondérée par la valeur du seuil. Les informations présentées en entrée d'un réseau de neurones et possédant des caractéristiques semblables vont s'accumuler jusqu'à ce que le seuil préfixé du réseau soit dépassé, l'amenant vers la décharge (cf. figure 2.). Cette notion de franchissement de seuils selon les niveaux d'activité permet de faire à partir des connexions de neurones logiques des systèmes de traitement des informations où toutes les opérations logiques peuvent être effectuées, mais avec la particularité remarquable d'intégrer une dynamique temporelle. Par exemple, les opérations logiques et /ou peuvent parfaitement être effectuées par des réseaux de quelques neurones. Le temps et la logique ne sont plus exclusifs car la résolution d'un problème logique nécessite intrinsèquement une durée de résolution. La figure suivante montre la formalisation mathématique de l'opération effectuée par un neurone formel. Un réseau de neurones est constitué de l'interconnexion d'un nombre important de ces unités de calcul élémentaire.

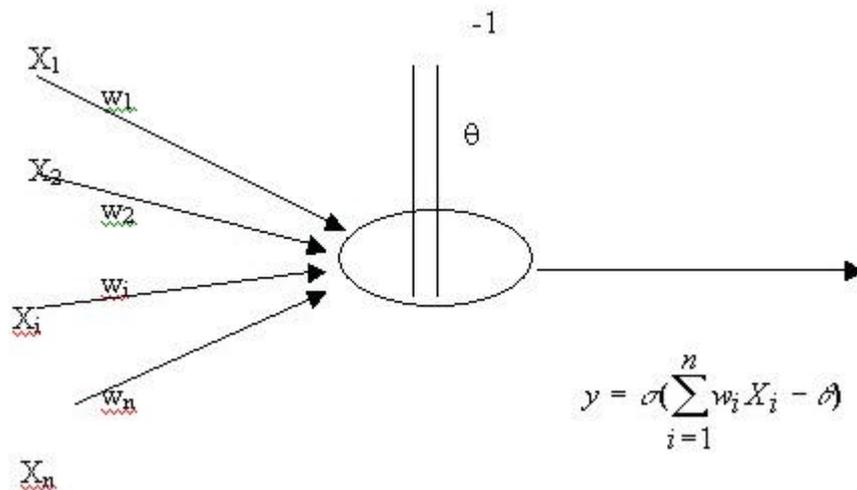


Figure 2. Opération de calcul de base d'un neurone formel

X_i : le signal transmis par le i -ème neurone en aval, w_i : le poids de la liaison synaptique entre le neurone et son i -ème parent, θ seuil, y le signal en sortie du neurone, σ la fonction d'activation. Le neurone est lié à n neurones afférents (X_1, \dots, X_n) au moyen des liaisons présynaptiques. Les signaux transmis par chacun, pondérés par le poids de la liaison w_i , s'additionnent. Si la somme dépasse le seuil, alors le neurone transmet un signal de sortie (potentiel d'action dans un neurone biologique).

1.3 Importance de la fonction de seuil

Chaque neurone formel possède des caractéristiques propres, en particulier un seuil de déclenchement, assimilable à un poids synaptique dont le dépassement implique la décharge du neurone, c'est-à-dire la transmission d'une information de sortie. La fonction seuil interne à chaque neurone formel assure que la valeur de sommation des potentiels pré-synaptiques dépassera pas certaines limites raisonnables (en général l'intervalle 0-1). Elle interdit les évolutions catastrophiques (effets de boucle où les valeurs deviennent de plus en plus grandes). Le Perceptron originel, inventé par Frank Rosenblatt en 1958, possédait une telle fonction de seuil (sorties toujours ramenées à 0 ou 1). Minsky et Papert ont démontré en 1969 les limites de cette sorte de perceptron (incapacité de discriminer des classes non linéairement séparables) induisant à attribuer ces limites à tous les réseaux de neurones. Cependant, les neurones à « sigmoïde » (la fonction douce et progressive) peuvent bien tout modéliser, même avec une seule couche. Enfin, dans le modèle probabiliste, la valeur de sortie du neurone y est une variable aléatoire.

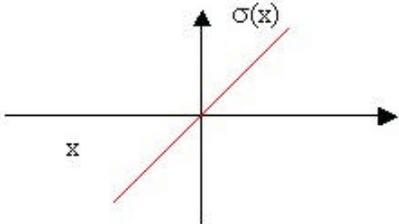
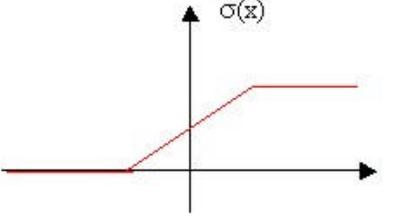
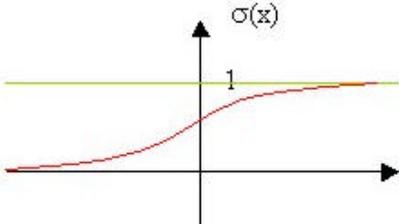
<p>Dans ce modèle, la fonction d'activation σ est la fonction identité. S'il a l'avantage d'être simple, il n'a que peu de rapport avec la réalité (le signal de sortie est non borné, linéaire par rapport aux signaux d'entrée, ce qui ne correspond pas du tout au fonctionnement des neurones biologiques).</p>	
<p>Le modèle linéaire par morceaux est plus réaliste en faisant apparaître des phénomènes de saturation.</p>	
<p>Le modèle sigmoïde est le modèle le plus utilisé dans la pratique</p>	

Tableau 1. Différentes fonctions seuil paramétrables dans les réseaux de neurones

1.4 Structure général d'un réseau de neurones

Comme les neurones formels sont interconnectés entre eux, leur niveau d'interconnexion définit des couches. On distingue une couche d'entrée contenant les neurones formels qui vont recevoir les informations primitives, puis après un certain nombre de couches intermédiaires dites couches cachées, une couche de sortie qui contient les neurones finaux transmettant les informations de sortie traitées par la totalité du réseau. Au contraire des systèmes de calcul symbolique qui codent les informations dans des symboles pour les calculer de façon séquentielle, les systèmes connexionnistes traitent les informations en parallèle. L'ensemble du réseau et des couches de neurones formels est alors modifié par les informations en entrée. Suivant les phases d'apprentissage, consistant à présenter les stimuli d'entrée un grand nombre de fois, les poids synaptiques des neurones formels sont modifiés. Il se définit alors petit à petit dans les couches cachées du réseau un attracteur vers lequel convergent l'ensemble des valeurs des poids synaptiques qu'on assimile alors à des trajectoires.

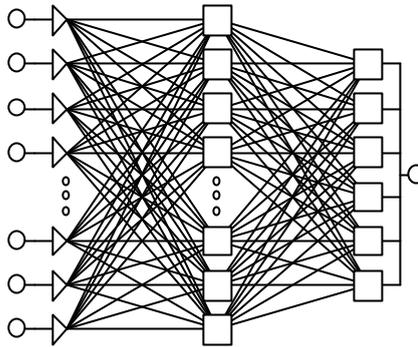


Figure 3. Schéma d'un réseau de neurones formels

On distingue la couche d'entrée (ronds), deux couches intermédiaires et une couche de sortie comprenant un seul neurone. Ce type de réseau est un perceptron et il peut être utilisé pour catégoriser un nombre important de variables d'entrée (une par neurone d'entrée).

1.5 Les mécanismes d'apprentissage

L'apprentissage d'un réseau de neurones artificiels est induit par une procédure itérative d'ajustement de ses paramètres internes (poids synaptiques et nombres de neurones). Cette procédure d'ajustement est décrite par un algorithme d'apprentissage. Celui-ci détermine alors le comportement du réseau. Ainsi le comportement d'un même réseau diffère selon l'algorithme d'apprentissage utilisé pour modifier ses paramètres. Il existe deux grandes stratégies d'apprentissage.

Apprentissage supervisé

La première est l'apprentissage supervisé. En déterminant des variables de sortie, on fait apprendre au réseau par itérations successives à modifier ses paramètres internes de façon à reconnaître ces variables. En quelque sorte, on propose au réseau des exemples (observations) auxquels sont associées des réponses désirées. Dans ce type d'apprentissage, la sortie produite par le réseau en réponse à une observation donnée est comparée à la réponse désirée (variable de sortie). La différence entre la réponse désirée et la réponse du réseau est alors utilisée pour adapter les paramètres du réseau de façon à corriger son comportement. Ce processus est répété de façon itérative jusqu'à obtenir le meilleur comportement. On présente en entrée chaque observation au réseau de neurones (c'est-à-dire qu'on met les valeurs décrivant le dossier dans les entrées), puis on propage ces valeurs vers la sortie. La prédiction que l'on obtient en sortie est (très probablement, surtout au début) erronée. On calcule la valeur de l'erreur (c'est donc la différence entre les valeurs désirées et les valeurs prédites), puis on « rétro-propage » cette erreur en remontant le réseau et en modifiant les poids proportionnellement à la contribution de chacun à l'erreur totale. On répète ce mécanisme pour chaque exemple des

données et tant que le taux de reclassement sur les données de test s'améliore. Ce mécanisme porte le nom de rétro-propagation de l'erreur. Il consiste à suivre la ligne de plus grande pente de la surface d'erreur, qui doit logiquement conduire à un minimum (local ou global). Sa capacité de convergence peut varier selon les problèmes posés, l'architecture du réseau, les fonctions d'activation retenues, et bien entendu, selon la base d'apprentissage et les variables sélectionnées en entrée du réseau. L'intérêt du Perceptron multi-couches est qu'il se prête idéalement à ce type d'algorithme de par sa structure et ses effets de propagation fortement structurés et maîtrisés. Le principe de l'algorithme est de remonter couche par couche, des neurones de sortie vers les neurones d'entrées et de modifier les poids synaptiques en amont de chaque couche, de manière à diminuer l'erreur commise en sortie. Le processus est en général itératif (le problème étant généralement non linéaire) : à chaque itération, l'erreur globale diminue. Signalons qu'il s'agit d'une technique générale (méthode de gradient) applicable à d'autres objets qu'à des réseaux de neurones. Par exemple on peut calculer une matrice de composantes principales de cette manière par ajustements successifs.

Apprentissage non supervisé

L'apprentissage non supervisé est une technique différente où on ne détermine pas de variables de sortie. Le réseau va de lui même catégoriser les variables d'entrée. Cet apprentissage est destiné à l'élaboration d'une représentation interne de l'espace des données d'entrée en identifiant la structure statistique sous-jacente des variables d'entrée sous une forme plus simple ou plus explicite. Ce type d'apprentissage est utilisé dans les réseaux compétitifs et dans les cartes de Kohonen qui seront décrits plus loin (cf. architecture des principaux réseaux de neurones).

En pratique, la création d'un réseau de neurones nécessite le formatage des données sous une forme qui soit reconnaissable par la couche superficielle du réseau (neurones d'entrée). Puis on choisit le type de réseaux que l'on veut utiliser. On présente alors les données en entrée en modifiant petit à petit les poids synaptiques et les autres paramètres afin de parvenir au but recherché, si l'on est en apprentissage supervisé. On laisse par contre le réseau évoluer tout seul vers son état stable, si l'on est en apprentissage non supervisé.

1.6 Architecture des principaux réseaux de neurones

Perceptron. Premier en date des réseaux de neurones, ce système comporte trois couches dont deux effectuent des traitements et une, la rétine, assure la réception des signaux. Le système fonctionne suivant une loi de convergence qui renforce l'activation d'un neurone si celui-ci est activé. Au cours du temps, cette règle de convergence aboutit à la création d'attracteurs dans la couche moyenne du réseau. Particulièrement efficaces pour le traitement des données spatiales, ces systèmes permettent de reconnaître des formes approximatives et de les comparer à des patterns appris, codés de façon délocalisée dans le réseau, permettant le fonctionnement d'une mémoire associative. Si un stimulus n'est pas trop éloigné de la forme apprise durant la phase d'apprentissage, alors il sera reconnu comme étant cette forme. Le problème de la catégorisation des connaissances est

particulièrement bien résolu par ce type de système qui accepte l'incertitude et la vectorise sous forme d'écart au centre de l'attracteur.

Adaline est un autre type de réseau spécialisé dans le traitement du signal fonctionnant à partir d'une règle d'apprentissage. Cette règle permet de réduire l'erreur observée entre l'état du réseau et la valeur correcte attendue par une modification des connexions des neurones de façon à la minimiser (par un algorithme dit de descente de gradient).

L'associateur linéaire est un modèle de réseau à deux couches avec une fonction d'activation linéaire utilisant la règle de Hebb. D. Hebb a proposé en 1949 l'idée que le cerveau s'adapte à son environnement en modifiant l'efficacité des connexions entre ses neurones. Le principe de Hebb postule qu'une synapse améliore son efficacité seulement quand l'activité des deux neurones qu'elle relie est corrélée.

Les réseaux récurrents permettent une interconnectivité quasi totale entre les neurones des différentes couches, y compris éventuellement en se bouclant sur eux-mêmes. Ils permettent la résolution de problèmes qui ne peuvent être résolus de façon algorithmique.

Le réseau de Hopfield inspiré des modèles physiques des verres de spin est constitué d'éléments bistables à connexions symétriques évoluant spontanément vers une réduction de l'énergie totale du système. Associé avec la règle de Hebb, ce réseau peut apprendre à mémoriser les exemples présentés en entrée sous la forme d'états stables. En phase d'exploitation, les stimuli présentés en entrée évolueront dans le réseau vers l'état stable le plus ressemblant. Un réseau de Hopfield fonctionne ainsi comme un véritable classificateur à mémoire associative. Cependant il existe une marge non négligeable d'erreurs due à l'existence d'états métastables, parfois très proches des états stables correspondant à une classification correcte.

Les réseaux compétitifs sont des réseaux où chaque neurone d'entrée est relié à chaque neurone de sortie et chaque neurone de sortie inhibe tous les autres et s'auto excite. Cette architecture génère une compétition inter-neurones aboutissant à ce que le réseau a tendance à reproduire l'organisation topographique des formes d'entrée. En d'autres termes, si l'on présente à ce type de réseau des objets quelconques, le réseau va reproduire dans ses états internes ses traits structuraux. Les réseaux compétitifs reproduisent une particularité du fonctionnement biologique des neurones, à savoir l'inhibition latérale. On sait en effet que lorsqu'un neurone biologique est excité, il transmet son excitation aux neurones voisins dans un rayon très court et inhibe par contre les neurones situés à plus grande distance.

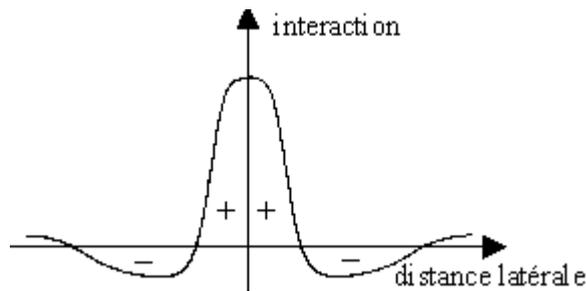


Figure 4.

Phénomène d'inhibition latérale dans les neurones biologiques et formels

Un neurone activé inhibe ses voisins. Il en résulte un phénomène de compétition interneurones dont le résultat est l'extraction des classes constituantes. Biologiquement, ce phénomène est à la source des champs récepteurs en neurophysiologie sensorielle.

Le degré d'interaction latérale est donc une fonction de la distance. Ces interactions latérales engendrent des réponses de groupes de neurones répartis autour du maximum local d'excitation. Les algorithmes d'apprentissage dans les réseaux compétitifs sont des simplifications algorithmiques utilisant l'idée de réponses localisées et d'interactions latérales. La forme la plus simple d'apprentissage compétitif modifie le vecteur poids du meilleur neurone à chaque étape de l'apprentissage.

Les cartes de Kohonen sont des réseaux de neurones compétitifs qui ont la particularité de réaliser une représentation ordonnée des données selon une topologie fixée, généralement une grille bi-dimensionnelle. En général, l'algorithme permet d'assurer que les données projetées dans un même voisinage sont des données proches. En revanche, il peut se produire que des données proches soient projetées dans des zones non voisines de la grille de représentation. Ces réseaux de neurones sont considérés comme des « classificateurs universels ». Ils résolvent notamment des problèmes de modélisation ou de classification linéaire. Le problème résolu devient un problème de régression linéaire multiple. La carte de Kohonen est une matrice de neurones représentant la couche de sortie. A chaque neurone est associé un vecteur de poids. Ces poids sont d'abord initialisés aléatoirement. L'algorithme d'apprentissage de ce réseau est itératif: on prend un vecteur de l'ensemble et on le « montre » au réseau. En comparant ce vecteur avec les poids des neurones, le neurone gagnant (le plus proche) est choisi. Cette modification est appliquée aussi pour les neurones voisins du neurone gagnant. Ce voisinage décroît avec le temps. Grâce à cette notion de voisinage, à la fin de l'algorithme (après avoir fait un certain nombre d'itérations) la carte montre une certaine organisation: les neurones proches ont des poids similaires. Il existe donc une préservation topologique entre l'espace d'entrée et l'espace de sortie. Quand le déroulement de l'algorithme se termine, les poids des neurones contiennent les vecteurs représentatifs de l'ensemble de vecteurs utilisé. Il s'agit d'un algorithme de recherche de centroïdes dans un espace vectoriel. L'auto organisation de la carte est conséquence de l'application de la notion de voisinage.

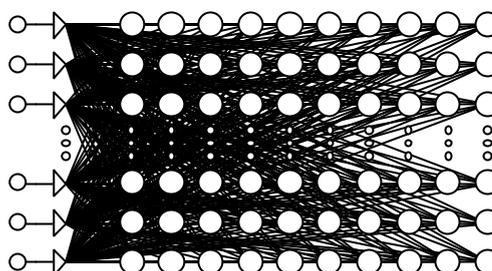


Figure 5. Réseau de type Kohonen.

Ce type de réseau présente une couche d'entrée et une couche de sortie constituée ici d'une matrice bidimensionnelle de neurones. Chaque neurone de cette couche est en compétition avec ses voisins.

De façon générale et pour l'ensemble de ces réseaux, la présentation de stimuli en entrée aboutit progressivement à la génération de noyaux denses dans les couches profondes. Ces noyaux denses correspondent aux traits structuraux des objets présentés en entrée. En phase d'exploitation, les formes présentées en entrée de réseau sont alors attirées vers les attracteurs et ainsi reconnues par leur activation, à la condition que les coordonnées du stimulus ne soient pas trop éloignées de celles acquises par apprentissage et qu'elles restent incluses dans le bassin d'attraction de chaque attracteur. Si elles sortent du bassin, alors elles sont attirées vers un attracteur voisin qui sera activé (changement d'état) et le processus s'apparentera à une catégorisation. L'ajout de couches intermédiaires et l'augmentation du nombre de neurones (et donc l'augmentation du nombre de connexions synaptiques) favorise les effets combinatoires. Les fonctions d'activations non linéaires et bornées contournent dans une certaine mesure les éventuels problèmes de multi-colinéarité liés à l'interdépendance des variables.

En pratique, le choix du type de réseau dépend de l'objectif que l'on veut atteindre. Si l'on souhaite faire des prédictions ou des modélisations de comportement d'un système, les réseaux de type Perceptron suffisent. Par contre, si l'on veut extraire de l'information de données ou réaliser des classifications, alors les réseaux compétitifs sont indiqués. La plupart des grands logiciels statistiques proposent les différents types de réseaux et permettent parfois de construire des réseaux hybrides. Dans tous les cas, les logiciels de statistique proposent un réseau standard, après le choix fait par l'utilisateur du type de réseau. Il convient ensuite de remanier le réseau en modifiant les fonctions seuils, le nombre de couches et donc la configuration globale du réseau. Ensuite, il faut sélectionner un algorithme d'apprentissage et choisir le nombre d'itérations du réseau. Pour toutes ces opérations, il faut procéder par essais / erreurs et donc de façon empirique. On sait que l'on a obtenu un bon réseau capable de traiter les données d'entrée et de réaliser des prédictions de façon optimale lorsque on atteint le phénomène de sur-apprentissage. A un moment donné, les taux obtenus par le réseau chutent brusquement après un dernier ajustement. La configuration précédente du réseau est alors généralement la configuration optimale.

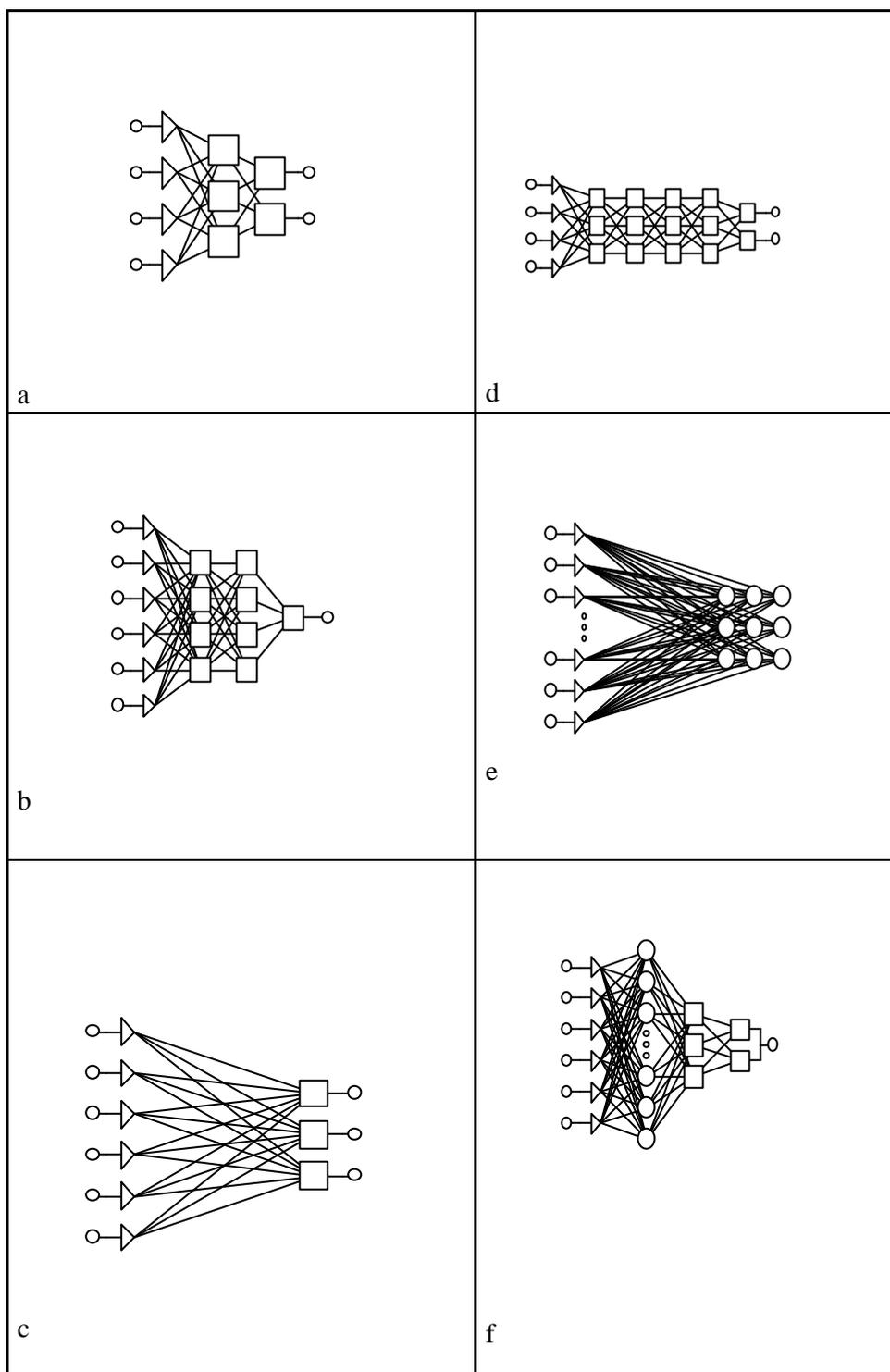


Tableau 2.
Principaux type de réseaux de neurones
(a) (b) (d) Perceptron multicouches (c) linéaire (e) Kohonen (f) radial

2. Applications en psychométrie

2.1 Avantages et inconvénients des réseaux de neurones

On peut donc définir un réseau de neurones artificiels comme un processeur parallèle de traitement d'informations distribuées présentant une propension naturelle à la mémorisation et à l'exploitation de connaissances acquises par l'expérience. Sa structure repose sur une interconnexion massive de cellules élémentaires de traitement d'information, les neurones formels, dont la représentation est un graphe dirigé. En cours d'apprentissage, un réseau peut être amené à modifier sa structure en modifiant les liaisons entre neurones. La connaissance acquise par un réseau de neurones est encodée par des forces ou des intensités évolutives des connexions établies entre les neurones formels. La forces des connexions liant les neurones réfèrent aux valeurs des poids ou coefficients qui définissent les paramètres du modèle interne du réseau. L'intérêt des réseaux de neurones artificiels réside dans le parallélisme de leur structure, leur capacité d'adaptation ainsi que dans leur mémoire distribuée. Ces propriétés sont à la source de la capacité de généralisation d'un réseau et donc de son aptitude à adopter un comportement correct (interpolation et extrapolation) en réponse à des variables d'entrée qui n'ont pas été rencontrés lors de la phase d'apprentissage.

Vus sur un plan mathématiques, les réseaux de neurones sont en effet des « approximateurs universels ». (théorème de Kolmogorov, 1957). Leur utilisation permet de passer directement des données au prédicteur, sans intermédiaire, sans recodage, sans discrétisation, sans simplification ou interprétation sujette à caution. Un réseau de neurones possède également une grande résistance au bruit ou au manque de fiabilité des données. L'idée d'apprentissage est plus simple à comprendre que les complexités des statistiques multi-variables. Elle est intuitive. Les réseaux de neurones sont également capables d'analyser des relations spatiales et topologiques. Enfin, le modèle fourni se prête bien à la visualisation, à une recherche de profils-type. Ces propriétés des réseaux de neurones sont intéressantes car elles sont complémentaires des techniques statistiques classiques qui imposent des fortes restrictions. Les régressions se font sur des fonctions de dépendance simples (linéaire, logarithmique) qui ne sont pas toujours très réalistes. On suppose souvent qu'il y a une seule fonction de dépendance sur l'ensemble de la population et non plusieurs sur des « niches » distinctes. On est obligé également de faire des hypothèses (distributions normales, équiprobabilités, variables non corrélées) dont il est probable qu'elles sont simplement approchées, mais qui sont nécessaires pour faire fonctionner ces méthodes. Enfin le traitement des calculs impose de faire appel à des méthodes compliquées pour transformer les données initiales. Par contre, un réseau de neurones bien conçu est capable de représenter n'importe quelle dépendance fonctionnelle et d'extraire de l'information sans modèle préconçu. A contrario, les réseaux de neurones présentent des inconvénients. Un réseau de neurones ne dispense pas de bien connaître son problème, de définir ses classes avec pertinence, de ne pas oublier de variables importantes, etc. Enfin, un réseau de neurones est une « boîte noire » qui n'explique pas ses décisions. Les réseaux de neurones ont une très bonne prédiction statistique (ayant la capacité de s'accommoder de valeurs très bruitées ou même

manquantes), mais ils sont complètement impossibles à inspecter. La perte partielle de compréhension est compensée par la qualité des prédictions.

2.2 Indications des réseaux de neurones en psychométrie

Du fait de leurs propriétés, l'application d'un réseau de neurones formels est indiquée dans les situations suivantes :

1. Le modèle du problème considéré est inconnu ou difficile à formaliser alors que l'on dispose d'un ensemble d'observations composées de variables d'entrée auxquelles sont associées des variables de sortie sans que l'on connaisse précisément la fonction de transfert qui les relie.
2. Les données du problème et en particulier les données d'entrée sont entachées de bruit lié à des erreurs de mesure ou à des données manquantes. Dans ce cas, la propriété de robustesse du réseau permet de pallier le bruit et d'exploiter néanmoins ces données.
3. On ne connaît pas la topologie des liens liant des variables descriptives mais on désire obtenir une connaissance sur la nature de cette topologie. La propriété auto-organisationnelle d'un réseau récurrent muni d'un apprentissage non supervisé par un algorithme de Kohonen permet d'obtenir une visualisation de la topologie des liens entre les variables d'entrée.
4. Le problème nécessite un traitement en temps réel capable de se modifier au fur et à mesure de l'évolution des données d'entrée.

Les trois premières situations sont fréquemment rencontrées en psychométrie et justifient l'application des réseaux de neurones dans ce domaine d'application. En effet, la construction d'un test (T1) psychométrique ou d'un questionnaire nécessite leur étalonnage sur une population de sujets qui ont été sélectionnés selon des critères d'homogénéisation ou de représentativité. Ces échantillons de sujets, construits pour les étalonnages, peuvent être l'objet d'une mesure psychométrique avec un autre instrument (T2) que celui du premier test dont on effectue l'étalonnage. Il est alors utile de disposer d'une méthode permettant de prédire l'appartenance à la catégorie de T2 à partir des réponses aux items du test T1. L'étude présentée dans la section suivante traite de cette question.

3. Etude sur la prédiction d'une catégorie externe

3.1 L'échelle BDI-II

L'inventaire de dépression de Beck révisé (BDI-II) est un questionnaire d'auto évaluation destiné à mesurer la sévérité de la dépression chez des sujets à partir de 16 ans. Il est composé de 21 items. Chaque item propose une question auquel le sujet répond en inscrivant sa réponse sur une échelle entre 0 et 3. L'inventaire a été conçu pour évaluer les symptômes dépressifs correspondant aux critères diagnostiques des troubles dépressifs selon le DSM-IV (version française, 1996). Ces critères sont les suivants :

1. Humeur dépressive
2. Diminution d'intérêt ou de plaisir
3. Perte ou gain de poids et/ou de l'appétit
4. Insomnie ou hypersomnie
5. Agitation ou ralentissement
6. Fatigue-perte d'énergie
7. Dévalorisation-culpabilité
8. Pensée-Concentration
9. Pensée de mort –suicide
10. Pas d'existence d'épisode mixte
11. Souffrance-baisse de fonctionnement cliniquement significatif
12. Non dus à des drogues ou à des médicaments
13. Non dus à un deuil

Le développement de la version française a été réalisé à partir d'un échantillon de 437 sujets (127 hommes et 310 femmes) de moyenne d'âge de 42 ans avec un écart type de 14.5. Parmi ces 437 sujets, 96 sont des sujets témoins et les 341 autres sont des sujets présentant divers types de pathologie, définis selon les critères du DSM-IV. Pour tous les sujets ayant répondu au BDI-II, le psychiatre ou le psychologue indiquait le diagnostic clinique actuel et remplissait une feuille de renseignements. Pour chaque sujet de l'échantillon clinique, il précisait le diagnostic clinique actuel suivant les critères du DSM-IV. Pour les sujets dont le diagnostic était dépression majeure, il notait alors les critères de diagnostic de dépression majeure, les critères de sévérité et le sous type clinique. Le diagnostic porté par le clinicien comporte les catégories suivantes :

1. Episode dépressif majeur
2. Double dépression
3. Dysthymie
4. Trouble anxieux isolé
5. Patient ambulatoire tout venant
6. Sujet témoin

Pour les sujets appartenant à la catégorie 1 – épisode dépressif majeur – le clinicien définissait les critères diagnostics :

1. Humeur dépressive
2. Diminution d'intérêt ou de plaisir
3. Perte ou gain de poids et ou de l'appétit
4. Insomnie ou hypersomnie

5. Agitation ou ralentissement
6. Fatigue et perte d'énergie
7. Dévalorisation-culpabilité
8. Pensée concentration
9. Pensée de mort-suicide
10. Pas de symptôme d'épisode mixte
11. Souffrance-baisse de fonctionnement significatif
12. Dépression non due à des drogues ou médicaments
13. Dépression non due à un deuil

Les critères de sévérité pour ces même sujets présentent une dépression majeure sont les suivants :

1. Dépression légère
2. Dépression moyenne
3. Dépression sévère sans caractéristiques psychotiques
4. Dépression sévère avec caractéristiques psychotiques congruentes
5. Dépression sévère avec caractéristiques psychotiques non congruentes

Enfin le sous-type clinique de ces sujets présentant une dépression majeure était également indiqué :

1. Trouble dépressif majeur épisode isolé
2. Trouble dépressif majeur récurrent
3. Episode dépressif actuel de trouble bipolaire I (un ou plusieurs épisodes maniaques avec un ou plusieurs épisodes dépressifs)
4. Episode dépressif actuel de trouble bipolaire II (épisodes dépressifs avec un ou plusieurs épisodes hypomaniaques).

Toutes ces catégorisations sont portées par le clinicien sur ces critères cliniques externes au BDI-II. Il est intéressant de vérifier si un réseau de neurones est capable de prédire qualitativement l'appartenance d'un sujet à ces catégories uniquement à partir des réponses aux items du BDI-II

3.2 Méthodologie générale

Nous avons utilisé le module de réseaux de neurones du logiciel *Statistica* (société Stat Soft). Ce module est indépendant du logiciel complet de statistique. Il peut être interfacé à *Excel* et à d'autres logiciels de statistique. Le principe d'utilisation est le suivant. On édite le tableau à double entrée des données d'un échantillon de travail. Chaque ligne correspond à une observation (un sujet) et chaque colonne correspond à des variables descriptives. Celles-ci comportent les items mais aussi les caractéristiques des sujets et également d'autres variables qui peuvent être associées aux observations mais qui sont indépendantes des items. La première tâche consiste à bien préparer les données en étant attentif à la nature symbolique ou numérique des données. Ensuite on attribue à chaque variable la fonction d'être soit une variable d'entrée (elle correspondra alors à un neurone de la couche d'entrée), soit une variable de sortie (elle correspondra alors à un neurone de la couche de sortie), soit d'être ignorée par le réseau. On effectue alors un tri au hasard dans cet échantillon de façon à générer deux sous-échantillons. L'un servira à la phase d'apprentissage, l'autre à la phase de vérification. On débute ensuite la construction du réseau en choisissant le type de réseau parmi la liste des réseaux

disponibles (Cf. le paragraphe architecture des réseaux). Le choix dépend des objectifs et de la structure des données. Pour les études présentées ici, nous avons opté pour la construction d'un réseau de type perceptron multicouches. Ensuite, il faut paramétrer le réseau en définissant les poids synaptiques des différentes couches. En pratique, on peut se contenter des valeurs par défaut que l'on peut modifier ensuite si le réseau se révèle inefficace. Après avoir sélectionné le nombre de cycle d'apprentissage (nous avons opté pour une valeur de 400 par défaut qui s'avérera suffisante dans la plupart des cas), on exécute le programme. Celui-ci présente les données d'entrée devant la couche d'entrée à chaque cycle d'apprentissage. On obtient ensuite une matrice de résultats présentant les taux de réussite des apprentissages sur les différentes variables de sortie. Généralement, ces taux sont faibles à la première expérience et il convient de rendre le réseau plus performant en appliquant soit une méthode de rétro-propagation des erreurs soit une méthode mathématique de descente par gradient. On obtient à la suite de cette deuxième application des résultats généralement meilleurs et on continue de la sorte jusqu'à obtenir les meilleurs taux. L'optimum est obtenu juste avant que ce produise le phénomène de sur-apprentissage qui aboutit à une chute catastrophique des taux d'identification. Il convient alors de sauvegarder le réseau ayant obtenu les meilleurs résultats sous la forme d'un fichier externe. Ensuite, on effectue un test de vérification en lançant le réseau sur un échantillon de sujets choisis au hasard n'ayant pas été inclus dans l'échantillon d'apprentissage. On obtient alors une nouvelle matrice de résultats correspondant aux performances réelles du réseau. Ce sont ces matrices qui sont présentées comme résultats numériques sur les tableaux et les figures présentés dans les résultats de la section suivante.

3.3 Résultats

Prédiction du diagnostic clinique

En réunissant dans une même catégorie, tous les patients déprimés pour les distinguer des troubles anxieux et des patients ambulatoires tout venant. Le réseau est un perceptron multicouches. Le nombre de cycles d'apprentissage est fixé à 400. On obtient une matrice de résultats présentés sous la forme d'un tableau présentant sur la dernière ligne les taux d'identification correcte (cf. tableaux 3 et 4, en gras). Le réseau est capable de différencier correctement les sujets témoins des autres sujets et d'identifier les sujets dépressifs et les sujets présentant d'autres types de pathologies. Par contre, il ne peut distinguer les sujets anxieux (0.50). En dissociant plus précisément les différentes catégories constituant de la catégorie générale de la dépression, nous obtenons des bons taux de reconnaissance des différentes catégories. Cependant tous ces taux sont soumis à des contraintes volumétriques. Plus l'échantillon d'entraînement est de taille réduite et moins le réseau est capable de réaliser de bonnes catégorisations (cf. tableau 5.).

	Témoin	Dépression	Anxieux	Patients ambulatoires tout venant
Total	96	190	42	109
Correct	86	186	21	105
erreurs	1	0	4	1
inconnu	9	4	17	3
Témoin	86	0	2	1
Dépression	0	186	0	0
Anx	0	0	21	0
PATV	1	0	2	105
	0.89	0.97	0.50	0.96

Tableau 3. Prédiction du diagnostic clinique

Perceptron multi-couches avec rétro propagation des erreurs 400 itérations
Ensemble de l'échantillon – tous types de dépression confondus

	Sujet témoin	Dépression majeure	Double Dépression	Dysthymie	Trouble anxieux	Patient ambulatoire tout venant
Total	96	119	31	40	42	109
Correct	85	101	18	20	17	86
erreurs	2	0	0	0	2	3
inconnus	9	18	13	20	23	20
TEMOIN	85	0	0	0	0	2
EPD	0	101	0	0	0	1
DD	0	0	18	0	0	0
DYSTH	0	0	0	20	0	0
TAXI	0	0	0	0	17	0
PATV	2	0	0	0	2	86
	0.88	0.84	0.58	0.50	0.40	0.78

Tableau 4.

Prédiction du diagnostic clinique sur l'ensemble des catégories
Perceptron multi-couches avec retro propagation des erreurs 400 itérations

	Sujet témoin	Dépression majeure	Double dépression	Dysthymie	Trouble anxieux	Patient ambulatoire tout venant
Total	39	69	18	20	19	52
Correct	23	40	4	5	3	26
Erreur	0	2	1	0	0	3
Inconnu	16	27	13	15	16	23
	0.58	0.57	0.22	0.25	0.15	0.5

Tableau 5.

Effet de volume sur les performances du réseau

Avec le même type réseau que précédemment mais avec un volume d'observations réduit et sélectionné au hasard, le réseau ne parvient pas à catégoriser correctement même en augmentant le nombre d'itérations (800) et en appliquant plusieurs types d'algorithme d'apprentissage.

Prédiction de l'appartenance nosologique

En utilisant un réseau du même type et en effectuant un nouvel apprentissage de reconnaissance des catégories nosologiques, on obtient également des taux acceptables de reconnaissances des catégories nosologiques à partir d'un échantillon de vérification (cf. tableau 6.).

	Non déprimé	TRM Récurrent	TRM isolé	Bipolaire 1	Bipolaire 2
Total	287	93	34	10	13
Correct	273	60	13	2	1
erreurs	1	1	2	1	0
inconnus	13	32	19	7	12
Non déprimé	273	1	2	1	0
TRMR	1	60	0	0	0
TRMI	0	0	13	0	0
BIPO1	0	0	0	2	0
BIPO2	0	0	0	0	1
	0.95	0.64	0.38	0.20	0.007

Tableau 6.

Prédiction de l'appartenance nosologique

Perceptron multi-couches avec retro-propagation d'erreurs. Le réseau catégorise les non déprimés des sujets déprimés mais distingue mal les sous-catégories internes au groupe des patients déprimés.

NORM sujet témoin, TRMI trouble dépressif majeur épisode isolé, TRMR trouble dépressif majeur récurrent, BIPO1 épisode dépressif actuel de trouble bipolaire 1, BIPO2 épisode dépressif actuel de trouble bipolaire 2.

Critères de sévérité

La BDI-II est surtout cliniquement utilisé pour dépister les cas de sévérité de la dépression. Il est intéressant de vérifier si un réseau de neurones est capable de prédire la sévérité d'une dépression en ne connaissant que les réponses aux items. Pour ce faire, on a utilisé l'étalonnage actuel et on a appris au réseau à reconnaître la sévérité de la dépression en la positionnant comme variable de sortie. Les résultats montrent une bonne capacité de prédiction pour deux des degrés de l'échelle de sévérité (cf. tableau 7.). Les autres degrés étant moins bien reconnus. Les sujets témoins sont pratiquement tous identifiés. Enfin, le réseau s'est révélé très sensible au sur apprentissage aboutissant à la dégradation rapide des résultats. Le réseau est un perceptron de 13 couches cachées. Le nombre de cycle d'apprentissage a été déterminé à 400. La méthode d'apprentissage est celle d'une descente de gradient.

Discussion

En conclusion, un réseau de neurones de type perceptron muni d'un algorithme de rétro propagation des erreurs est capable de prédire l'appartenance diagnostic des sujets sur la base des réponses aux items du BDI-II à la condition qu'il existe suffisamment de sujets dans chacune des catégories. De façon générale, on observe dans ces différents résultats que les taux de catégorisation sont liés au nombre de sujets dans chaque catégorie. Plus il existe de sujets dans une catégorie à l'intérieur de l'échantillon ayant été utilisé pour l'apprentissage et plus le réseau est capable de les distinguer des autres sujets appartenant à une autre catégorie. Tous ces résultats sont conformes aux attentes liées aux propriétés générales des réseaux de neurones.

Degrés de sévérité	A	D	C	B	E	F
Total	287	67	64	12	5	2
Correct	283	57	52	3	1	1
erreurs	0	0	5	2	0	0
inconnu	4	10	7	7	4	1
	0.98	0.85	0.81	0.25	0.20	0.50

Degrés de sévérité	A	D	C	B	E	F
Total	153	29	28	6	2	1
Correct	151	25	24	4	0	1
erreurs	0	0	2	0	1	0
inconnu	2	4	2	2	1	0
	0.98	0.86	0.85	0.66	0.00	1.00

Degrés de sévérité	A	D	C	B	E	F
Total	134	38	36	6	3	1
Correct	132	34	31	3	1	0
erreurs	0	1	1	1	0	0
inconnu	2	3	4	2	2	1
	0.98	0.89	0.86	0.50	0.33	0.00

Tableau 7.

Prédiction du critère de sévérité

Réseau perceptron multicouches 400 itérations rétro-propagation des erreurs. En haut sur l'ensemble de la population, au milieu et en bas sur deux échantillons choisis au hasard. Le réseau parvient à une prédiction correcte du niveau de sévérité.

A sujets non déprimés, B Dépression légère, C Dépression moyenne, D Dépression sévère sans caractéristiques psychotiques, E Dépression sévère avec caractéristiques psychotiques congruentes, F Dépression sévère avec caractéristiques psychotiques non congruentes

4. Etude sur la classification automatique

4.1 Utilisation de la carte topologique de Kohonen

Une des propriétés les plus remarquables des réseaux de neurones concernent les capacités de certains d'entre eux (les cartes topologiques de Kohonen) à expliciter les topologies implicites liant les dimensions des variables d'entrée. Une carte topologique est un procédé d'auto-organisation qui cherche à projeter des données dont la dimension est en général grande dans un espace de faible dimension. En fin d'apprentissage, la carte reproduit les corrélations présentes dans les données d'entrée. Les données initiales sont projetées sur un espace discrète et régulier de faible dimension (1D ou 2D). Grâce au procédé d'auto-organisation, la topologie qui lie les données initiales est conservée au niveau des réponses proposées par la carte.

A partir d'un réseau récurrent de type Kohonen, nous avons cherché à vérifier si ce type de réseau était capable de retrouver les dimensions des facteurs sous-jacentes aux variables d'entrée aux données du BDI-II. La première étape consiste à enlever les variables de sortie de façon à présenter au réseau uniquement les réponses aux items de façon à être en apprentissage non supervisé. Ensuite, on construit un réseau récurrent de type Kohonen muni d'une couche de sortie correspondant à une matrice bi-dimensionnel de neurones. On entraîne le réseau sur ces données pendant un certain nombre d'itérations. Puis on recherche sur la couche de sortie l'existence d'un éventuel phénomène de clusterisation aboutissant à ce que certains neurones de la couche de sortie présentent un taux d'activation différent des autres. On définit ainsi une sorte de topologie bidimensionnelle des activations sur la couche de sortie. On peut alors remonter jusqu'aux observations et mettre en évidence des regroupements d'observations présentant des profils de réponses aux items ayant été clusterisés dans le comportement du réseau et on peut alors les identifier comme catégories constituantes.

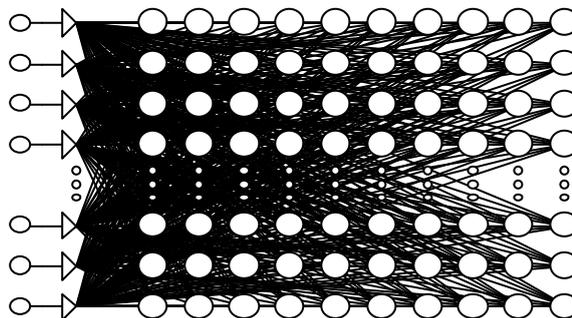


Figure 6.

Couche de sortie du réseau de type Kohonen montrant la matrice de neurones

La figure 7. montre les réponses de la couche de sortie d'un réseau de type Kohonen muni de 100 neurones disposés sur une matrice de 10 neurones de large. En présentant en entrée du réseau les réponses aux items du BDI-II, on observe un phénomène de clusterisation (en gris sur le tableau).

	1	2	3	4	5	6	7	8	9	10
1	3	7	1	6	6	13	3	5	2	1
2	0	2	4	5	4	1	2	2	1	0
3	1	10	7	2	1	7	3	3	2	4
4	10	6	10	2	4	3	4	2	0	6
5	3	1	3	4	2	1	3	2	0	2
6	2	1	0	6	3	1	0	0	1	0
7	6	0	2	7	6	0	1	1	2	0
8	2	3	3	29	18	1	3	2	0	0
9	4	9	16	0	3	10	3	4	5	4
10	15	17	44	16	2	0	1	0	2	6

Figure 7.
Matrice de résultats

Scores des neurones gagnants sur la couche de sortie en forme de matrice bi-dimensionnelle de taille 10x10 neurones. Les unités sont des grandeurs d'activation des neurones gagnants. On recherche les valeurs les plus élevées puis on remonte jusqu'aux observations.

La figure 8. présente ces mêmes données sous une forme graphique. On voit que la surface de réponse présente des pics correspondants à des activations des neurones de la couche de sortie. Ces pics correspondent à des facteurs internes sous-jacentes aux variables d'entrée (les réponses aux items) des sujets de l'échantillon considéré.

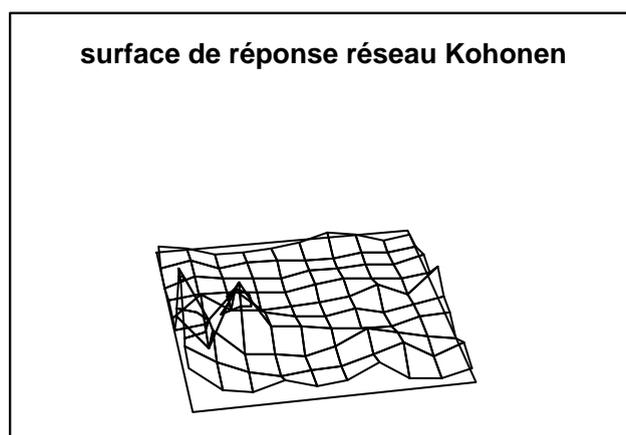
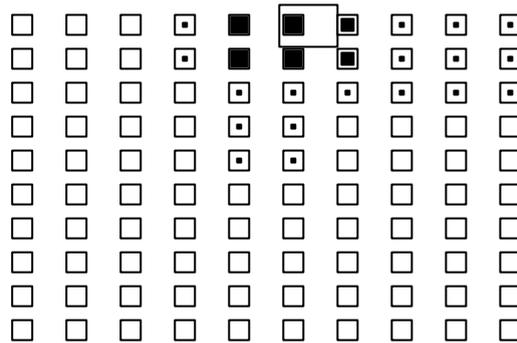


Figure 8.
Surface de réponse de la couche de sortie du réseau de type Kohonen
Les pics montrent des convergences d'activation.

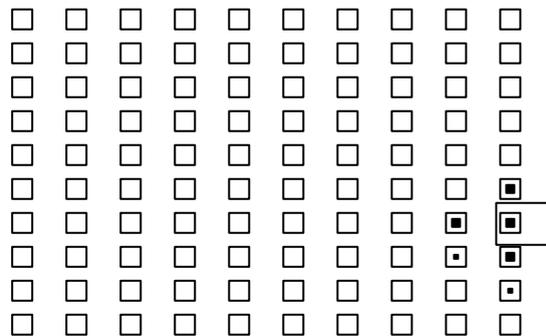
Chaque petit carré des matrices présentées sur les figures suivantes correspond à un neurone de la couche de sortie du réseau de Kohonen entraîné avec les données du BDI-II. Les carrés marqués par une marque noire de plus ou moins grande extension correspondent à des neurones activés. On observe un regroupement différent des neurones activés selon le type de cas. Par exemple, le cas N°167 montre des activations dans les neurones situés en haut et au milieu de la matrice. Ce cas est un sujet tout venant.



Cas N°167 sujet tout venant

Figure 9.

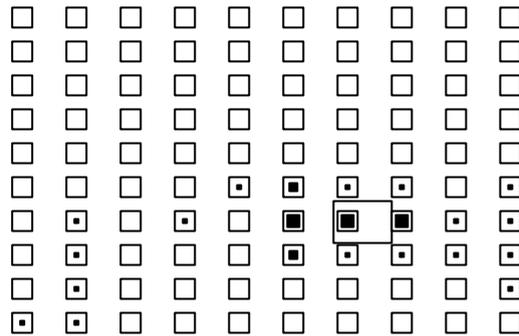
Par contre, le cas 17 montre une activation des neurones situés à un autre endroit de la matrice.



Cas N°17 sujet dépression majeure

Figure 10.

Le cas suivant montre également une activation d'autres neurones situés à une autre localisation de la couche de sortie du réseau de Kohonen. Ce cas s'avère être un sujet anxieux. Ainsi, les différentes catégories cliniques sont bien révélées sur la carte topologique bidimensionnelle de la couche de sortie du réseau de Kohonen.



cas N°133 Trouble anxieux

Figure 11.

Lorsque on effectue une superposition de l'ensemble des cas, on obtient une carte topologique globale où l'ensemble des catégories peut être positionné et identifié sur la matrice. La réalisation de cette carte topologique globale nécessite de remonter vers les observations à partir de la détection d'un cluster de neurones vainqueurs et d'aller rechercher dans les variables associées à quelle catégorie le sujet appartient. Puis on reporte sur la carte le nom de la catégorie. On obtient alors la carte topologique globale présentée sur la figure suivante.



Figure 12.

Carte topologique de Kohonen sur les données d'entrée de la BDI-II.

PATV Patients tout venant, DYSTH Dysthymie, EPD épisode de dépression majeure, DD double dépression, témoin Témoin, TAXI Trouble anxieux

On observe bien une disposition topologique des catégories sur le plan de la matrice. Des singularités peuvent être observables comme la duplication des zones correspondant à la dysthymie par exemple ou le regroupement local des sujets témoins et des patients tout venant. Ces éléments peuvent suggérer des interprétations ou des vérifications. Dans le cas de cette étude, nous étions en présence de données comportant l'information sur les catégories. En fait, cette méthode est plus intéressante lorsqu'on ignore les catégories ou les facteurs constituants et que l'on s'en sert comme un classificateur. Dans ce cas, la méthode est la même sauf qu'il faut attribuer des noms provisoires aux clusters rencontrés sur les matrices de neurones et ensuite remonter jusqu'aux observations. On essaye ensuite d'identifier les classes découvertes par le réseau.

4.2 Procédure standard en classification automatique

En résumé, la démarche à suivre pour la recherche de catégories ou de facteurs constituants est la suivante :

1. On prépare les données en attribuant à toutes variables le rang de variables d'entrée. L'absence de variables de sortie comme le choix du réseau impliquent donc un apprentissage non supervisé.
2. La deuxième étape consiste dans la construction du réseau de Kohonen et en particulier construire la matrice bi-dimensionnelle de la couche de sortie. Le choix de la taille et le nombre de neurones de cette couche sont deux paramètres importants. Rappelons qu'un réseau de Kohonen réalise une projection des dimensions de l'espace des variables d'entrée dans un espace de sortie aux dimensions plus réduites.
3. On exécute ensuite le réseau sur les données en itérant la passation de l'algorithme.
4. On visualise ensuite l'état d'activation des neurones de la couche au travers d'une matrice numérique ou d'une représentation graphique. On recherche alors l'existence des regroupements (neurones vainqueurs).
5. On attribue un nom ou un code provisoire à ces groupements de neurones sur la matrice. Le nombre des groupements de neurones indique le nombre de catégories présentes dans l'espace des données.
6. A partir de l'identification des neurones vainqueurs, on cherche les observations correspondantes à une activation maximale de ces neurones et on regroupe ces observations. Ces observations ont en commun l'activation de ces neurones vainqueurs. A partir des caractéristiques particulières de ces observations (autres données adjointes ou traits communs à toutes les observations), on peut alors inférer la nature des facteurs ou des catégories mises en évidence par la carte. Dans le cas où on ne dispose pas d'autres informations sur les observations ou qu'elles n'ont apparemment rien en commun, on peut

alors inférer l'existence de nouvelles catégories ou de nouveaux facteurs constituants.

5. Etude sur la prédiction de données manquantes

5.1 Le problème des données manquantes

Une des applications les plus intéressantes des réseaux de neurone en psychométrie concerne le problème des données manquantes. On sait ainsi que de nombreux questionnaires sont confrontés au problème de la non réponse. On distingue deux classes de non-réponse :

1° La non réponse totale lorsqu'un individu échantillonné ne fournit aucune réponse à l'ensemble du questionnaire.

2° La non réponse partielle lorsque l'individu ne répond pas à une partie plus ou moins importante du questionnaire.

Chaque type de non réponse nécessite une technique particulière de correction. La non-réponse totale est corrigée par des techniques de re-pondération qui consistent à augmenter le poids d'échantillonnage des répondants. On élimine ainsi les biais liés au comportement des non-répondants (il n'y a aucune raison pour que les comportements des répondants et des non-répondants soient semblables). Les techniques utilisées pour la correction des non-réponses partielles sont celles de **l'imputation**.

L'imputation consiste à remplacer chaque donnée manquante par une donnée estimée en fonction de renseignements obtenus pour le même individu et pour les individus proches. L'avantage de ces techniques est qu'une fois les données imputées, la matrice croisant les individus et les variables est complète et les données imputées peuvent alors être traitées comme des données réelles. Les principales méthodes d'imputation sont les suivantes:

1° Méthodes de déduction. Une valeur manquante est déduite à l'aide d'une règle logique construite à partir des réponses aux autres questions.

2° Méthodes de substitution. Une valeur manquante est remplacée par la valeur correspondante d'une observation similaire. Ce répondant similaire peut être choisi dans l'enquête courante (méthodes de hot deck qui donnent pour la valeur manquante la valeur observée d'un individu répondant, le donneur, choisi selon une procédure adéquate, soit aléatoire, soit séquentiel, soit métrique) ou dans d'autres sources (méthodes de cold deck qui utilisent l'information obtenue à partir des répondants d'une autre enquête).

3° Méthodes de prédiction. L'ensemble des données est utilisé pour construire un modèle qui prédit la valeur de l'item manquant à partir des valeurs des items corrélés (parmi les méthodes classiques de prédiction par imputation, on peut citer l'imputation par la moyenne, la régression...).

Les réseaux de neurones artificiels pour imputer des valeurs appartiennent à la troisième catégorie. Dans ce cas, une valeur manquante est remplacée par la sortie calculée par le réseau. Les autres valeurs corrélées décrivant l'observation sont fournies en entrée du réseau. Dans un premier temps, on apprend au réseau à reconnaître la variable de sortie. Le réseau apprend donc les liens implicites reliant l'ensemble des variables d'entrée à cette variable de sortie. Puis on effectue un test pour vérifier l'apprentissage en définissant au hasard un échantillon de la population des sujets. On vérifie ainsi que le réseau a bien codé les liens et qu'il est capable d'effectuer la reconnaissance. On en déduit alors que cette variable de sortie est bien liée aux autres variables d'entrée, qu'on peut retrouver sa valeur en cas d'absence en utilisant le codage interne au réseau.

Nous avons utilisé les données d'étalonnage du test R-CMAS Echelle révisée d'anxiété manifeste pour enfant, (Cecil R. Reynolds, Bert O. Richmond, adaptation française, Dana Castro). Ce questionnaire comporte 37 items sous forme de questions auquel l'enfant doit répondre par oui ou non. La cotation du questionnaire comporte 5 notes. La note totale d'anxiété est basée sur 28 items regroupés en trois sous-échelles d'anxiété : Anxiété physiologique, Inquiétude/hypersensibilité et préoccupations sociales/concentration. Les neuf items restants de la R-CMAS constituent l'échelle de Mensonge. La note brute de chaque sous-échelle représente le nombre total d'items pour lesquels le sujet a répondu oui à cette sous-échelle. Les notes brutes sont converties par étalonnages en notes standardisées et en percentiles.

Nous disposons d'un fichier comportant 1355 observations avec les réponses aux items du questionnaire. Pour tenter la restitution de données manquantes, nous avons opté pour la méthode suivante. Nous avons sélectionné un item et nous l'avons décrété variable de sortie (item 37). Tous les autres items ont été considérés comme des variables d'entrée et on a ignoré l'ensemble des autres variables (caractéristiques du sujet). Ensuite nous avons appris à un réseau Perceptron à reconnaître les valeurs de cette variable de sortie au travers d'une itération (600) d'un algorithme de rétro-propagation. Ensuite nous avons testé la capacité de ce réseau à prédire les données manquantes.

Comme chaque item peut avoir comme réponse oui ou non, la probabilité de prédire une réponse juste est de 0.5. En utilisant un réseau perceptron multicouches et un gradient de rétro-propagation, les résultats montrent une probabilité très nettement augmentée à 0.92 à la condition d'utiliser l'ensemble des données pour l'apprentissage. Dans ce cas, on peut utiliser ce réseau après son apprentissage pour inférer la réponse oui ou non d'un sujet n'ayant pas répondu à l'item 37 avec une probabilité faible d'erreur. On a procédé de même avec un autre item (33) appartenant à une autre échelle. Les résultats sont sensiblement comparables.

Item 37	Oui	non	Item 33	Oui	non
Total	776	579	Total	836	519
Correct	735	513	Correct	824	491
erreur	23	37	erreur	9	24
inconnu	18	29	inconnu	3	4
	0.94	0.88		0.98	0.94
	0.92			0.97	

Tableau 8.

Taux de reconnaissance des réponses au RCMAS
sur deux items à réponse binaire

Si l'on veut maintenant inférer des réponses manquantes automatiquement sur un nombre important de sujets, il faut connaître le comportement du réseau devant un groupe de sujets présentant des réponses manquantes. Par contre, le même réseau utilisé en coupant l'échantillon au hasard entre deux échantillons l'un d'apprentissage et l'autre de vérification donne des résultats décevants et inexploitable en pratique (cf. tableau suivant).

	Ensemble d'apprentissage		Ensemble de vérification	
	oui	non	oui	non
Total	387	291	389	288
correct	314	239	224	119
erreur	7	9	54	71
inconnu	66	43	111	98
oui	314	9	224	71
non	7	239	54	119
taux	0.81	0.82	0.57	0.41

Tableau 9.

Par contre, en changeant complètement de type de réseau et en remplaçant l'architecture d'un perceptron par celle d'un réseau de type radial, on obtient des résultats nettement plus satisfaisants. L'ensemble d'apprentissage a permis un comportement optimal du réseau et l'ensemble de vérification montre des taux améliorés par rapport au perceptron. Toutefois, on obtient pas des taux utilisables comme ceux réalisés par un perceptron entraîné sur l'ensemble des observations et testé ensuite par un cas unique.

	Ensemble d'apprentissage		Ensemble de vérification	
	oui	non	oui	non
Total	392	286	384	293
correct	392	286	257	200
erreur	0	0	127	93
inconnu	0	0	0	0
oui	392	0	257	93
non	0	286	127	200
taux	1.00	1.00	0.66	0.68

Tableau 10.

Réseau de type radial

L'utilisation d'un réseau de type radial se heurte aussi à une difficulté propre à ce type de réseau, à savoir un phénomène d'oscillation dynamique. Le réseau traite correctement tous les oui, mais pas les non, puis bifurque et réalise l'inverse. Tous ces résultats concernent un type particulier de réponse puisqu'il s'agit de réponses binaires oui / non. Il serait intéressant de tester la capacité d'un réseau à imputer des réponses aux items à variables numériques. Enfin, ce résultat pour être généralisé devrait être testé en croisant l'ensemble des items pour vérifier l'homogénéité de l'imputation.

5.2 Procédure d'imputation par réseau de neurones

En pratique, la méthode d'imputation des données manquantes par réseaux de neurones consiste à réaliser les étapes suivantes :

1. Vérifier l'adéquation de la méthode aux besoins. Cette méthode n'est applicable que dans le cas où l'on dispose de peu de sujets présentant des non-réponses à certains items alors que l'on veut disposer de données complètes avec un fort taux de certitude.
2. On sépare un échantillon présentant des sujets avec des données complètes.
3. On identifie l'item présentant des données manquantes et on le qualifie comme variables de sortie.
4. On construit un réseau de type perceptron auquel on apprend par apprentissage supervisé à reconnaître les valeurs de la variable de sortie. Cet apprentissage nécessite la plupart du temps l'utilisation d'une retro propagation des erreurs. Il convient de faire attention au sur apprentissage. Dans certains cas en effet, les taux chutent brutalement au cours d'une opération de retro propagation des erreurs. Il faut alors retrouver la configuration du réseau précédent qui devient alors le modèle optimal. On le sauvegarde alors soigneusement. On calcule alors le taux de réussite en

apprentissage. En entrée de ce même réseau, on présente ensuite les réponses aux items d'un sujet présentant une non-réponse sur l'item sélectionné (cf. le point 3). Le réseau prédit alors la réponse avec le taux de réussite calculé au point 4.

	Ensemble d'apprentissage		Ensemble de vérification	
	oui	non	oui	non
Total	383	295	393	284
correct	27	295	0	284
erreur	356	0	393	0
inconnu	0	0	0	0
oui	27	0	0	0
non	356	295	393	284
taux	0.07	1.00	0	1.00

	Ensemble d'apprentissage		Ensemble de vérification	
	oui	Non	oui	non
Total	383	295	393	284
correct	359	100	388	14
erreur	24	195	5	270
inconnu	0	0	0	0
oui	359	0	388	270
non	0	195	5	14
taux	0.93	0.33	0.98	0.04

Tableau 11.

Réseau de type radial montrant un comportement de bifurcation. Tous les Oui ou tous les Non peuvent à peu près prédits mais pas les deux ensembles.

6. Etudes en cours utilisant des réseaux de neurones

6.1 Etude longitudinale sur les implants cochléaires chez l'enfant

Cette étude en cours est réalisée sous l'égide du CTNRHI. Elle est commanditée par le ministère des affaires sociales et de la santé à la suite d'un avis du comité national d'éthique. Elle a pour objet l'évaluation de l'impact psychologique, linguistique et social des implantations cochléaires chez l'enfant sur dix ans. Les premiers résultats doivent être publiés dans le courant de l'année 2002. Un des principaux outils d'évaluation utilisés pour cette étude est le PSA, profil socio-affectif, édité par les ECPA. Le PSA a été passé avant et après implantation chez tous les enfants du groupe d'enfants implantés et du groupe témoin d'enfants sourds non implantés mais appariés en âge et sur le plan clinique. Le PSA est passé à intervalles réguliers (tous les 6 mois environ). On dispose ainsi de données évolutives au fil du temps. A côté de ces données recueillies au PSA, on dispose de nombreuses autres variables décrivant les dimensions linguistiques, audiométriques, audiophonologiques et sociales de l'enfant, ainsi que des variables techniques concernant les stratégies de réglage des implants cochléaires, les conditions de port prothétique.

Le problème posé est de vérifier si l'implantation génère une modification significative dans telle ou telle dimension. Les techniques statistiques classiques (t-test) sont a priori suffisantes à ceci près que le nombre de variables est très étendu et que l'on ne dispose de modèle théorique très précis pour tester statistiquement telle ou telle hypothèse. En effet, on ne sait pas précisément pourquoi dans tel cas une implantation cochléaire est associée à une amélioration sensible dans tel ou tel domaine. C'est justement sur ce point que l'utilisation des réseaux de neurones présente un intérêt. On les utilise de plusieurs manières différentes. Au travers d'un apprentissage non supervisé et de la constitution d'un réseau de Kohonen, on cherche à mettre en évidence dans la couche de sortie du réseau des topologies de liens entre les variables qui ne sont pas a priori connues ou suspectées. Dans ce cas, on attribue à l'ensemble des variables de l'étude le rang de variables d'entrée et on les traite avec un algorithme de Kohonen. On repère les clusters apparents sur la carte topologique et on remonte ensuite jusqu'aux observations. On interprète alors ces regroupements d'observations. La seconde utilisation des réseaux de neurones sur cette étude concerne la prédiction. On définit dans un premier temps un indice de réussite de l'implantation cochléaire. Cet indice est composite et intègre le point de vue du sujet lui-même, de ses parents, et des professionnels qui s'occupent de l'enfant (professeurs, orthophonistes). Ensuite, on construit un réseau de type Perceptron multicouches auquel on apprend à reconnaître l'indice de satisfaction placé comme variable de sortie (en apprentissage supervisé) à partir des données recueillies lors des bilans pré implants (PSA, données audiophonologiques, etc.). Le réseau ainsi entraîné encode les liens entre les données d'entrée et l'indice de satisfaction. Placé devant un nouveau cas (ou un groupe de cas), le réseau peut prédire l'indice de réussite. Celui-ci peut alors être utilisé, en association avec les avis cliniques, pour la prédiction de réussite d'une implantation cochléaire.

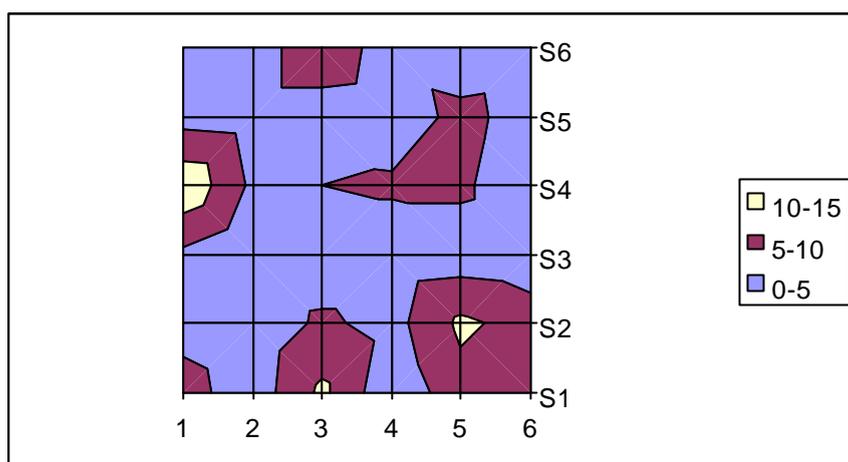


Figure 13.

Exemple de classification automatique par un réseau de Kohonen des résultats aux items du PSA sur 33 enfants implantés cochléaires. On observe sur la matrice (6 x 6) des neurones de sortie la constitution de clusters correspondants à des enfants présentant des profils qualitatifs de réponses particuliers. Ces enfants peuvent être ensuite identifiés. Les unités correspondent à des niveaux d'activation.

6.3 Evaluation des risques psychologiques chez les enfants de patients traumatisés crâniens.

Cette étude (2001-2002) est destinée à rechercher, à quantifier et enfin à qualifier les éventuelles difficultés psychologiques existantes chez des enfants dont l'un des parents a eu un traumatisme crânien. Cette étude est réalisée sous l'égide de l'ADEP cellule d'étude et de recherche sur la psychopathologie du traumatisme crânien, le service de rééducation neurologique de l'hôpital Raymond Poincaré à Garches, le Centre Médical Jacques Arnaud et le service de psycho-réhabilitation de Limoges. Pour des raisons de faisabilité, l'étude ne comporte pas de groupe témoin. Le test des ECPA, R-CMAS a été préconisé pour évaluer l'anxiété des enfants et en inférer des hypothèses psychopathologiques en association avec un questionnaire demi-directif comportant une série d'items destinés à son interprétation. Nous avons le projet d'analyser ces données avec un réseau de neurones à fin de tester d'une part les liens entre l'échantillon de sujets de l'étude et l'échantillon d'étalonnage pour lequel on a déjà construit un réseau. En cas de différences significatives entre les comportements du réseau devant les deux échantillons, nous espérons pouvoir identifier les variables impliquées. Enfin, une seconde utilisation de réseau concerne l'identification des liens entre les variables des réponses aux items du questionnaire semi-directif grâce aux propriétés de convergence d'un réseau de type Kohonen.

7. Conclusions

Bien que liminaires et demandant à être approfondies et complétées sur d'autres types de données, les études présentées ici montrent que les réseaux de neurones peuvent être utilisés avec profit dans l'analyse des données psychométriques. De nombreuses applications sont possibles. Nous avons vu que ces réseaux pouvaient être utilisés pour effectuer un test de cohérence d'une série d'items sélectionnés pour la qualité discriminante de tel ou tel facteur avec des catégories déterminées avec des instruments externes (étude BDI-II). Ils peuvent être aussi utilisés pour réaliser une prédiction d'appartenance d'un sujet à telle ou telle catégorie en utilisant le réseau comme modèle représentatif de la population d'étalonnage qui a servi à l'apprentissage. Le réseau fonctionne alors comme prédicteur à partir de la topologie qualitative des réponses et non de la sommation des notes aux items permettant ainsi une alternative à la détermination arithmétique des notes seuils.

En dehors de la prédiction pure de l'appartenance d'un groupe de sujet à telle ou telle catégorie, l'intérêt réside aussi dans l'évaluation possible de l'homogénéité des échantillons. En effet si un échantillon a permis à un réseau de catégoriser correctement les réponses aux items d'un premier test en prédisant l'appartenance à une variable d'un deuxième test alors on peut poser qu'un deuxième échantillon distinct du premier, même de même taille, testé par le même réseau aura sensiblement les mêmes résultats s'il est globalement apparié. Par contre, si le deuxième échantillon testé par le même réseau ne parvient pas à des résultats de prédiction correcte, alors on peut en conclure qu'il ne peut être apparié au premier échantillon. Le même raisonnement peut être utilisé pour tester la taille nécessaire d'un échantillon d'observation. En testant progressivement un réseau avec un échantillon de taille croissante, on pourrait par essais /erreurs déterminer la taille minimale pour que la catégorisation s'effectue de façon correcte.

Les réseaux de neurones permettent de traiter les variables simultanément et de conserver une cohérence entre elles. Du fait des propriétés de convergence vers l'attracteur du centre du réseau, des modifications ou des altérations des données présentées en entrée n'ont pas d'incidences sur la qualité de reconnaissance du réseau à la condition que ces modifications restent incluses dans certaines limites. Ainsi, non seulement les réseaux peuvent garder une bonne efficacité en présence de données manquantes, mais ils peuvent également, moyennant un apprentissage préalable sur des échantillons complets, permettre la prédiction de données manquantes sur d'autres échantillons. Toutefois, nous n'avons testé cette fonction sur des variables binaires. Une étude complémentaire sur des variables numériques serait nécessaire.

Enfin, les propriétés de classification automatique des données permettent de faire des recherches sur les classes constituantes, sans modèle préconçu, grâce aux réseaux de type Kohonen. Ces réseaux permettent de restituer dans la couche de sortie les topologies cachées des liens entre les variables d'entrée.

Les réseaux de neurones sont ainsi appelés à venir compléter les méthodes statistiques conventionnelles utilisées en psychométrie. En particulier, ils sont appelés à jouer un rôle utile dans les domaines où les données sont fortement bruitées par la présence de nombreuses données manquantes et à chaque fois que

l'on ne dispose pas d'un modèle théorique expliquant les liens entre les variables internes d'un test ou entre ces variables et celles d'un autre test. Cependant, il convient de rappeler que les réseaux de neurones ne peuvent expliciter par eux-même la façon dont ils opèrent. Leur capacité de prédiction est doublée d'une impossibilité à expliciter comment une prédiction se réalise. C'est une limite forte à l'usage exclusif des réseaux de neurones qui doivent trouver leur place en psychométrie en complément des autres méthodes statistiques.

Bibliographie

Amit D.J. Modeling Brain Function - The World of Attractor Neural Networks , Cambridge, Cambridge University Press, 1989.

BDI-II Inventaire de dépression, Aaron T. Beck, Robert A. Steer, Gregory K. Brown, Editions du Centre de Psychologie Appliquée.

Caron N. Les principales techniques de correction de la non-réponse et les modèles associés, INSEE, « Méthodologie Statistique » N°9604

Wodzisaw D (coll.) *Neural-based classification and visualization methods with applications to psychometry*. 34th International Seminar on Statistics and Clinical Practice, Warsaw 24-28.06.1996 (extended abstract), pp. 93-94

Hebb D. The Organisation of Behavior , New-York, Wiley, 1949.

Jodouin J.F «Présentation des modèles connexionnistes» , Intellectica, 9/10, 1990.

Meunier B. «La physique des réseaux neuronaux», Intellectica, 9/10, 1990.

Minsky M., Papert S. Perceptrons , Cambridge, Mass, MIT, Press, 1969.

R-CMAS Echelle révisée d'anxiété manifeste pour enfant, Cecil R. Reynolds, Bert O. Richmond, adaptation française Dana Castro, Editions du Centre de Psychologie Appliquée.

Rumelhart D.E, G. E. Hinton, and J. L. McClelland *Parallel distributed processing: Explorations in the microstructure of cognition*, Bradford Books, 1986

Smolensky P. «Représentation in Connectionist Networks» , Intellectica, 9/10, 1990.

Remerciements

Louis-Charles Vannier pour la prise en main logicielle et autres bonnes idées...