# ALTERNATIVE NEURONALE EN TARIFICATION SANTE

Jean-Marc AOUIZERATE [1]
*Gras Savoye* [2]

**Résumé** :

Cet article propose une alternative aux méthodes traditionnelles de tarification frais de santé, conçue sur la base d'un modèle de *Réseau de Neurones Artificiels*. Cette technique permet de capter automatiquement des dépendances non linéaires de haut niveau entre les variables explicatives dans le but d'affiner le tarif. L'autre spécificité du modèle présenté est son approche globale du tarif qui n'est plus estimé pour chaque poste de dépenses comme dans le cadre classique.

Le modèle a été bâti sur un portefeuille de plus de 600 000 bénéficiaires en testant différentes variantes. Le réseau offrant les meilleurs résultats a ensuite été confronté avec succès à d'autres modèles avancés. Enfin, dans des conditions réelles d'utilisation, les tarifs prédits par le modèle neuronal se sont avérés être plus proches de la consommation réelle qu'avec des outils de tarification standard dans 79% des cas.

**Mots-clefs :** Réseaux de Neurones - Perceptron Multicouche - Tarification - Assurance Frais de Santé

**Abstract:**

This paper proposes an alternative to traditional pricing methods for additional health insurance. The mathematical model used is an *Artificial Neural Network*. This technique can automatically capture high-level nonlinear dependencies between explanatory variables in order to refine the price. Another specificity of the model is the global approach of the price. Contrary to the classical techniques, the price is not estimated for each medical expenses item.

The model was built on a portfolio of more than 600 000 beneficiaries by testing different variants. The network which provides the best results was confronted successfully with other advanced models. Finally, in real conditions, the prices predicted by the neural model were closer to real consumption than standard pricing tools in 79% of cases.

**Keywords:** Neural Networks - Multilayer Perceptron - Pricing - Health Insurance

[1] Jean-Marc Aouizerate est actuaire en prévoyance et santé collective. Contact : aouizerate.jeanmarc@neuf.fr
[2] Ces travaux ont été effectués de 2008 à 2010 au sein du Département Prévoyance et Retraite de Gras Savoye.

## 1. INTRODUCTION

Avec une progression de 6% en 2008, le « marché » de la complémentaire santé représentait un chiffre d'affaires de 29 milliards d'euros selon les déclarations des organismes complémentaires au Fonds de financement de la Couverture Maladie Universelle. Ce marché connait toujours une progression soutenue. De 2001 à 2008, il a progressé de 61% soit une croissance annuelle moyenne de 7,4%.

L'environnement de la santé en France évolue inexorablement vers une plus grande sollicitation des complémentaires santé. Les budgets alloués à la couverture maladie représentent une part de plus en plus importante pour les entreprises pour lesquelles la santé devient un réel enjeu économique. Le pilotage des régimes prend ainsi une dimension majeure au sein des sociétés qui souhaitent mieux maîtriser leurs dépenses tout en maintenant un niveau de protection convenable pour leurs salariés.

Le contrôle des coûts, la mise en place de réformes, l'apparition de garanties innovantes ont donné une impulsion à l'activité tarification chez les différents acteurs de l'assurance et du courtage. Dans l'optique de mieux répondre aux attentes des entreprises, une réflexion sur le pilotage des régimes frais de santé a été menée. Celle-ci a abouti à l'élaboration d'une technique novatrice permettant d'appréhender plus précisément le risque et par conséquent le coût de leur régime.

Sur de nombreux postes médicaux tels que l'optique ou le dentaire, les comportements des consommateurs sont étroitement liés aux niveaux de couverture dont ils disposent. De manière générale, la part de financement supportée directement par le salarié peut freiner sa consommation et par conséquent limiter les remboursements effectués par sa complémentaire. Cette influence du reste à charge intervient à différents niveaux. Premièrement, un effet dissuasif peut se produire, entraînant une diminution du nombre d'actes consommés. Deuxièmement, les assurés, davantage responsabilisés par leur contribution pécuniaire, peuvent être plus attentifs aux prix pratiqués, ce qui implique une baisse des coûts moyens. Troisièmement, les remboursements assureur peuvent être optimisés en ajustant les dépenses engagées sur les limitations contractuelles des garanties. Or, dans le cadre d'une méthode actuarielle de tarification classique d'un régime complémentaire santé, ces phénomènes ne sont pas toujours recensés : pour une population donnée, les fréquences et distributions des coûts sont souvent supposées constantes. Cette constatation a conduit à expérimenter ici une modélisation tarifaire qui intègre ces corrélations. Dans cette logique, la consommation médicale pouvait s'inscrire assez

naturellement dans une approche neuronale et constituer ainsi une alternative aux méthodes plus conventionnelles de cotation.

Les techniques de modélisation non linéaires par apprentissage, développées récemment, suscitent beaucoup d'intérêt tant au sein du milieu académique que dans de nombreux secteurs d'activité tels que l'industrie, les sociétés de services ou encore les laboratoires de recherche. Les réseaux de neurones, inspirés directement du schéma de traitement de l'information des neurones biologiques, ont bénéficié du succès de la notion sous-jacente d'intelligence artificielle. Ces outils mathématiques performants permettent de classifier ou bien de régresser des fonctions complexes sans modélisation analytique particulière (aspect « boite noire »). Ce concept novateur, associé à l'évolution de la vitesse des traitements informatiques, ont fortement contribué à l'essor de ce modèle.

Dans un premier temps, une ébauche des concepts de base du modèle de réseau de neurones sera présentée dans le cadre général du Perceptron Multicouche.

Ensuite, un modèle de réseau de neurones sera bâti dans le cadre de la tarification d'un régime santé. Différentes techniques seront testées puis comparées pour sélectionner au final l'architecture et le mode d'apprentissage offrant les meilleurs résultats.

Enfin, les résultats obtenus par le réseau de neurones sélectionné seront d'abord comparés à d'autres méthodes avancées, puis leur cohérence avec la logique économique de l'assurance santé sera vérifiée. En conclusion, le modèle sera mis en pratique dans un contexte réel d'utilisation en le confrontant à une méthode classique.

## 2. LE MODELE NEURONAL

### 2.1 Introduction

Les modèles de réseaux de neurones artificiels s'inspirent directement du mécanisme biologique de pensée du cerveau humain en cherchant à imiter son processus d'apprentissage. L'architecture d'un réseau est composée de plusieurs neurones reliés entre eux par des connections et organisés en couches successives. L'information se propage au travers du modèle de neurone en neurone, de la couche d'entrée vers celle de sortie en passant par les couches intermédiaires (cachées). Le modèle peut ainsi être entraîné pour une tâche spécifique en ajustant ses paramètres (poids) au fur et à mesure que des exemples lui sont présentés.

Le neurone formel est apparu en 1943 [CUL43] avec une première modélisation mathématique du neurone biologique. Les premières règles d'apprentissage remontent à 1949 avec celle de Hebb [HEB49]. Puis, le premier réseau de neurones apparaît en 1958 avec le Perceptron de Rosenblatt [ROS58]. En 1986, l'algorithme d'apprentissage de rétropropagation , proposé initialement par Werbos en 1974 [WER74], est rendu populaire par Rumelhart [RUM86]. Ce n'est qu'à partir des années 1990 que ce modèle rencontra un succès notamment en raison de ses capacités d'apprentissage et de généralisation.

Un des avantages de ce modèle est de pouvoir traiter des cas non linéaires. Il trouve son utilisation dans le cadre de modélisation de fonctions, prédiction, classification, reconnaissance de formes. Les domaines d'application sont assez vastes : industrie, finance, télécommunications, informatique, environnement, assurance automobile [BOU07], [DUG03], [KIT02], [LOW96] [MUL06], [PAG11], [PEL98], [PEL99], [SMI02], [YEO05], plus rarement en assurance santé [HSU08], [LIN08], [LIO08], [ORT06], [YEH10], etc.

Pour approfondir le sujet de manière plus détaillée, le lecteur pourra se référer aux ouvrages suivants : « *Apprentissage statistique* » de Gérard Dreyfus [DRE08], « *Data mining et statistique décisionnelle* » de Stéphane Tufféry [TUF10] ou encore « *The Elements of Statistical Learning* » de Trevor Hastie [HAS08].

## 2.2 Neurone formel

Un neurone formel est une modélisation mathématique du neurone biologique. Il est généralement constitué de plusieurs entrées et d'une sortie par analogie aux dendrites et à l'axone. Les grands principes y sont repris, notamment la sommation des entrées. A chacune de ces entrées, est associé un poids qui correspond aux actions excitatrices et inhibitrices des synapses. Une phase d'apprentissage est alors utilisée pour ajuster ces coefficients.

Schématiquement, le neurone calcule la somme pondérée des entrées qu'il perçoit auquel se rajoute un seuil. Ensuite, il applique à la valeur obtenue une fonction dite d' « activation » classiquement non linéaire. Enfin, la valeur finale est affectée en sortie du neurone.

Considérons le cas d'un neurone formel à $n$ entrées auxquelles sont affectées $n$ grandeurs numériques notées $x_1$ à $x_n$ , pour définir la règle de calcul.

Un poids synaptique noté $w_1$ à $w_n$ est associé à chaque entrée, la sommation pondérée est obtenue de la façon suivante :

$$w_1.x_1 + ... + w_n.x_n = \sum_{i=1}^{n} w_i.x_i$$

Le neurone commence à devenir actif lorsque la sommation pondérée dépasse un seuil (ou biais) noté $w_0$. Il est additionné à cette grandeur, pour ensuite être transformé par la fonction d'activation notée $\phi$ :

$$\phi\left( w_0 + \sum_{i=1}^{n} w_i.x_i \right)$$

En général, l'expression de la valeur de sortie est simplifiée en ajoutant une entrée fictive $x_0$ fixée à la valeur 1 :

$$\phi\left( \sum_{i=0}^{n} w_i.x_i \right)$$

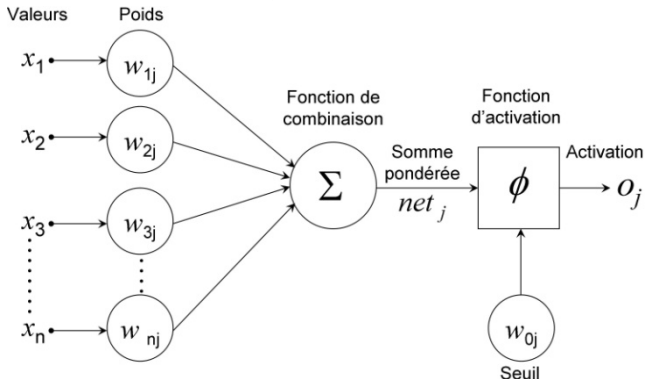



*Figure 1: Neurone formel*

### 2.3 Fonction d'activation

Le choix de la fonction d'activation est important pour obtenir un modèle utile en pratique car il influe sur les propriétés du neurone formel. De préférence strictement croissante et bornée, une fonction d'activation se doit de respecter une certaine forme de régularité et des conditions de dérivabilité.

Les plus fréquemment utilisées sont des sigmoïdes ayant une forme de S et symétrique par rapport à l'origine comme la fonction logistique, la tangente hyperbolique ou l'arc-tangente. Il existe cependant d'autres fonctions : le pas unitaire, la linéaire seuillée, la gaussienne ou encore l'identité.

### 2.4 Propagation de l'information

Un réseau de neurones artificiels est composé d'un ensemble de neurones formels qui, associés en couches, fonctionnent en parallèle. Chacune des couches effectue un traitement indépendamment des autres pour transférer le résultat de son analyse à la couche suivante. L'information se propage de couche en couche, de l'entrée vers la sortie. Le nombre de couches cachées est variable selon la nature du modèle. En fonction du sens et de l'orientation des différentes connexions définissant la structure du réseau, la valeur obtenue en sortie du neurone par la fonction d'activation est alors transmise comme entrée aux neurones suivants qui lui sont reliés. Par ce mécanisme, l'information peut alors circuler de neurone en neurone et de couche en couche au travers du réseau. Les couches dites d'entrée et de sortie constituent l'interface avec l'extérieur. La couche d'entrée reçoit les variables en entrée du modèle et la couche de sortie renvoie les résultats attendus. Les neurones des autres couches, internes au réseau, sont appelés neurones cachés. Conventionnellement, les neurones d'entrée ont une fonction d'activation identité pour ne pas modifier l'information. En fonction de l'algorithme d'apprentissage, l'information peut également être propagée en arrière (*back propagation*). De manière générale, chaque neurone, à l'exception de ceux des couches d'entrée et de sortie, est connecté à tous les neurones de la couche précédente et de la couche suivante.

Un réseau est caractérisé par trois composantes dont les deux premières définissent l'architecture :

- le type d'interconnexion,
- le choix des fonctions d'activation,
- le mode d'apprentissage (ou comment estimer les poids).

### 2.5 Le Perceptron Multicouche

Ce modèle intègre une ou plusieurs couches cachées entre l'entrée et la sortie. A l'exception des couches d'entrée et de sortie, chaque neurone est connecté avec tous ceux de la couche suivante et tous ceux de la précédente. Par contre, il n'existe pas de liaison entre des neurones issus de la même couche. Les fonctions d'activation principalement utilisées sont des fonctions à seuil ou des sigmoïdes. Comparativement au Perceptron monocouche, il a l'avantage de résoudre des problèmes non linéairement séparables ainsi que des problématiques logiques plus complexes comme le OU exclusif. La phase d'apprentissage est supervisée et régie par la règle de correction de l'erreur.

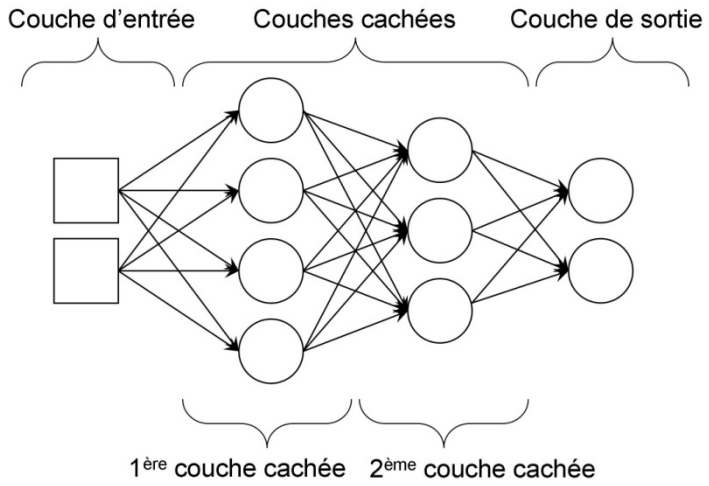



*Figure 2: Exemple d'architecture de PMC*

### 2.6 Les autres types de réseaux

Le modèle de base est le Perceptron Multicouche mais d'autres types de réseaux existent :

- les réseaux R.B.F. (Radial Basic Functions) : dans cette spécialisation du P.M.C., la fonction d'activation dans leur couche cachée est une gaussienne. Contrairement aux sigmoïdes, leur zone d'activité est limitée dans l'espace pour leur permettre de travailler localement dans l'espace des entrées (cf. [BOR07] p. 12, 13, 31-34, [DRE08] p. 78, 79, 190, [HAS08] p. 212-214, [TUF10] p. 220-223),
- les réseaux récurrents tels que les réseaux compétitifs (réseaux de Kohonen) et les cartes auto organisatrices : les corrélations et les régularités présentes dans les données en entrée sont détectées pour adapter les réponses en conséquence afin d'apprendre à catégoriser des vecteurs d'entrée (cf. [BOR07] p. 77-90, [DRE08] p. 349-426, [TUF10] p. 223-227).

### 2.7 L'algorithme de rétropropagation du gradient de l'erreur

Cet algorithme itératif a pour objectif de faire converger le modèle vers un minimum d'erreur (appelée aussi fonction coût) entre les valeurs attendues et celles calculées. Pour chaque neurone du réseau, le gradient de l'erreur est calculé afin de mesurer la contribution de chacun des poids synaptiques à l'erreur commise. Par ce biais, les corrections sont

effectuées au fur et à mesure que les exemples d'apprentissage sont présentés. Comme la modification d'un poids influe sur tous ceux des neurones des couches suivantes, les corrections d'erreurs doivent être propagées de la dernière couche vers la première.

Définissons la fonction d'activation comme étant une sigmoïde de la manière suivante : $\phi(x) = \frac{1}{1+e^{-x}}$

Celle-ci est une approximation indéfiniment dérivable de la fonction à seuil de Heaviside dont la dérivée est relativement simple à calculer : $\phi(x)' = \frac{e^x}{(1+e^x)^2} = \phi(x)(1-\phi(x))$

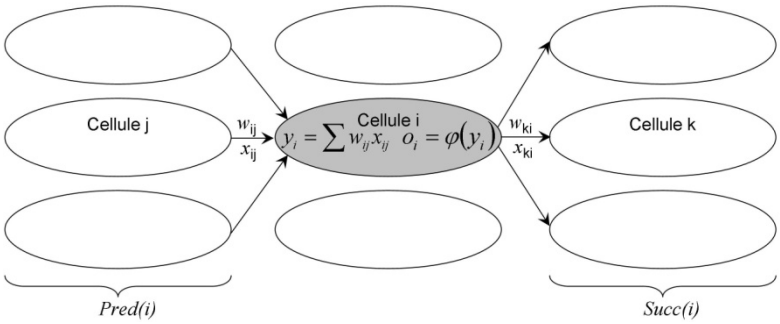



*Figure 3: Notations utilisées*

La difficulté majeure réside dans la complexité des notations définies de la manière suivante :

- l'échantillon d'apprentissage comprend $S$ exemples
- chaque exemple d'apprentissage est noté par un indice $s$
- chaque cellule est définie par un indice noté $i$
- le réseau comprend $p$ cellules de sortie
- la sortie attendue pour une cellule de sortie est notée $c_i$
- le poids synaptique de la cellule $j$ vers la cellule $i$ est noté $w_{ij}$
- l'ensemble des cellules en entrée de la cellule $i$ est noté $Pred(i)$ (prédécesseur)
- l'ensemble des cellules en sortie de la cellule $i$ est noté $Succ(i)$ (successeur)
- l'entrée totale de la cellule $i$ est notée $y_i$ de telle sorte que : $y_i = \sum_{j \in Pred(i)} w_{ij} x_{ij}$

- la sortie de la cellule $i$ est notée $o_i$ de telle sorte que : $o_i = \phi(y_i)$
- l'erreur d'apprentissage est notée $E$
- le nombre d'entrées est noté $n$
- le nombre de sorties est noté $p$
- le vecteur des poids synaptiques associé à tous les liens du réseau est noté $\vec{w}$

L'erreur commise sur un exemple est notée :

$$E(w) = \frac{1}{2}\sum_{k=1}^{p}\left(c_k - o_k\right)^2$$

Les poids $w_{ij}$ sont mis à jour selon la méthode de descente de gradient :

$$\Delta w_{ij} = -\frac{\partial E}{\partial w_{ij}}\varepsilon = -\underbrace{\frac{\partial E}{\partial y_i}}_{\delta_i} x_{ij}\varepsilon = \delta_i x_{ij}\varepsilon$$

$y_i$ est la somme pondérée de toutes les entrées, $\delta_i$ le gradient local du neurone et $\varepsilon$ le pas d'apprentissage.

Supposons comme fonction d'activation de la couche cachée une fonction logistique et une fonction linéaire pour la couche de sortie.

Le calcul de $\frac{\partial E}{\partial w_{ij}}$ est effectué avec la technique de rétro-propagation du gradient :

Si la cellule $i$ est de sortie :

$$\delta_i = -\frac{\partial E}{\partial y_i} = c_i - o_i$$

Si la cellule $i$ est interne :

$$\delta_i = -\frac{\partial E}{\partial y_i} = o_i\left(1 - o_i\right) \times \sum_{k \in succ(i)} \delta_k \times w_{ki}$$

L'algorithme peut alors se résumer en trois phases :

- la propagation avant,
- le calcul des $\delta$ (gradient locaux),
- la modification des poids.

L'algorithme continue jusqu'à ce qu'un critère d'arrêt soit satisfait.

## 2.8 Early stopping

Cette technique permet d'anticiper le phénomène de sur-apprentissage en arrêtant prématurément la présentation des exemples. Un premier cycle est réalisé en présentant

aléatoirement tous les exemples de la base d'apprentissage. Un second est réalisé sur la base de validation. Le processus s'arrête dès lors que l'erreur simulée sur la base de validation a atteint son minimum.

Soient le nombre d'exemples de la base d'apprentissage $N$, la k-ième valeur attendue $y_k$ et la k-ième valeur prédite $f(x_k)$, l'Erreur Quadratique Moyenne (E.Q.M.) peut être définie comme suit :

$$EQM(i) = \frac{1}{N}\sum_{k=1}^{N}\left(y_k - f(x_k)\right)^2$$

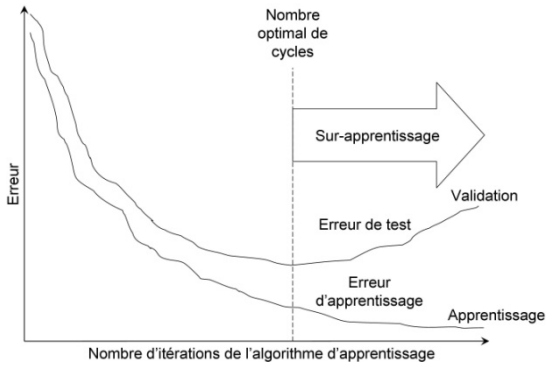



*Figure 4 : Technique du early stopping*

Une bonne pratique consiste à séparer la base de données en trois parties, à condition bien sûr que la taille le permette :

- une base d'entraînement,
- une base de validation,
- une base de test.

La base d'entraînement sert à l'apprentissage des coefficients, celle de validation est utilisée pour déterminer le nombre d'itérations optimal et celle de test permet de vérifier les performances du réseau ainsi que sa capacité de généralisation. Une répartition entre ces trois bases peut être 60% pour l'apprentissage, 20% pour la validation et 20% pour les tests.

## 3. MISE EN ŒUVRE DU MODELE DE TARIFICATION SANTE

### 3.1 Postulat du modèle

A l'origine de cette réflexion sur le choix d'un nouveau modèle, quelques observations montraient que les méthodes traditionnelles n'utilisaient pas suffisamment les corrélations entre les différents facteurs.

- **La part des dépenses restant à la charge du bénéficiaire peut constituer un réel frein modifiant ainsi son comportement.** Dans ce cas, la fréquence de consommation de même que les montants de frais réels sont d'autant plus faibles que la contribution financière supportée par le salarié est importante. Cette influence peut être plus ou moins prononcée selon la Catégorie Socio Professionnelle pour des motivations purement économiques.
- **Les bénéficiaires et les professionnels de santé peuvent être parfois tentés d'optimiser les remboursements complémentaires en fonction des garanties dont ils disposent.** Sur l'optique et le dentaire, il n'est pas rare de constater un ajustement des prix sur les limitations de garanties ou bien un transfert de budget des montures sur les verres.
- **Certains postes de dépenses, souvent consommés simultanément, peuvent être étroitement liés.** Par exemple, un contrat ne proposant qu'un faible remboursement des montures pourrait limiter la consommation sur ce poste et entraîner indirectement une diminution de la consommation de verres.
- **L'aspect préventif d'une garantie peut aussi limiter la consommation future sur un autre poste.** Par exemple, la mise en place de garanties couvrant des soins dentaires tels que la parodontologie peut à terme diminuer la consommation de couronnes, même principe avec la chirurgie de l'œil et les équipements optiques.

Or, ces effets, qui ont un impact direct sur le coût du régime, ne sont pas intégrés dans les méthodes de tarification classiques. Ce constat amène à penser qu'elles pourraient être sensiblement améliorées en trouvant un modèle capable d'intercepter des liaisons de haut niveau entre les variables.

En l'occurrence, les solutions apportées par les réseaux de neurones permettent d'y répondre parfaitement et notamment grâce aux deux propriétés essentielles suivantes qu'ils possèdent :

La propriété d' « Approximateurs universels » : Un Perceptron Multicouche doté d'une unique couche cachée, avec un nombre suffisant de neurones, peut approximer n'importe quelle fonction avec la précision voulue [HOR93].

La propriété d' « Approximateurs parcimonieux » : Les réseaux à une couche cachée

forment une famille d'approximateurs parcimonieux, ce qui signifie qu'à nombre égal de paramètres, il est possible d'approximer correctement davantage de fonctions qu'avec des polynômes [BAR93].

## 3.2 Définition du modèle de base

Le modèle de référence a été bâti à partir de la consommation médicale observée en 2007 sur l'ensemble du portefeuille Gras Savoye (hors personnels expatriés), soit plus de 610 000 bénéficiaires. Ses caractéristiques sont indiquées ci-après. Compte tenu de la volumétrie importante des données, l'ensemble des traitements et calculs ont été réalisés sous le logiciel SAS et notamment grâce à la plateforme de Data Mining SAS Entreprise Miner [MAT05].

### *3.2.1 Le choix des variables*

Les informations qui ont été retenues en entrée du modèle sont indiquées par bénéficiaire ayant consommé on non sur une année. Les personnes n'ayant pas bénéficié de soins sont également prises en compte dans le coût du régime. Par souci d'homogénéité, la consommation annuelle est corrigée du temps de présence sur l'année.

Les variables jugées pertinentes qui ont été sélectionnées sont les suivantes :

- l'identifiant du bénéficiaire (N° interne)
- le sexe (variable binaire : Homme ou Femme)
- l'âge (variable continue)
- le type de bénéficiaire (variable catégorielle : Adhérent, Conjoint ou Enfant)
- le collège (variable catégorielle : Ensemble, Cadre, ETAM, Non cadre)
- l'activité (variable binaire : actif ou inactif)
- le choix, type de contrat (variable catégorielle : Obligatoire, Base, Option)
- le régime (variable catégorielle : Général, Alsace-Moselle)
- le secteur d'activité (variable catégorielle : 24 secteurs recensés)
- la région (variable catégorielle : 22 régions recensées)
- les indicateurs de garanties (21 variables continues)
- le remboursement moyen (variable continue) qui correspond à la variable cible du modèle

### *3.2.2 L'indicateur de garanties*

L'indicateur utilisé ici permet de mesurer la performance d'une garantie de manière

homogène quel que soit son expression (% de la B.R., % du P.M.S.S., euros, ...). Il correspond au taux de couverture fictif qu'aurait la garantie d'un contrat si elle était appliquée à la consommation de l'ensemble du portefeuille. La prise en compte du portefeuille comme référentiel neutralise les écarts de prix des actes qui influent directement sur le calcul du taux de couverture [AOU10]. Il présente également l'avantage de pouvoir être estimé directement à partir de la consommation médicale sans avoir à recourir à la lecture des garanties.

Pour un contrat et un poste donnés, il est obtenu à l'issue des trois étapes suivantes :

- discrétiser la répartition des dépenses moyennes par acte en tranches (de 50 €, par exemple),
- calculer le taux de couverture sur chaque tranche (remboursements rapportés aux dépenses totales),
- pondérer les tranches selon la répartition du portefeuille (et non plus celle du contrat étudié).

### *3.2.3 L'architecture*

L'architecture du modèle de base retenue est un Perceptron Multicouche comportant une seule couche cachée et 20 neurones cachés. Entre deux couches, tous les neurones sont connectés sans aucune liaison directe. La fonction d'activation sigmoïde est de type tangente hyperbolique.

Le schéma suivant illustre l'architecture du modèle où $\omega$ symbolise les poids, $\sum$ la sommation, $\int$ la fonction d'activation sigmoïde et $/$ celle linéaire :

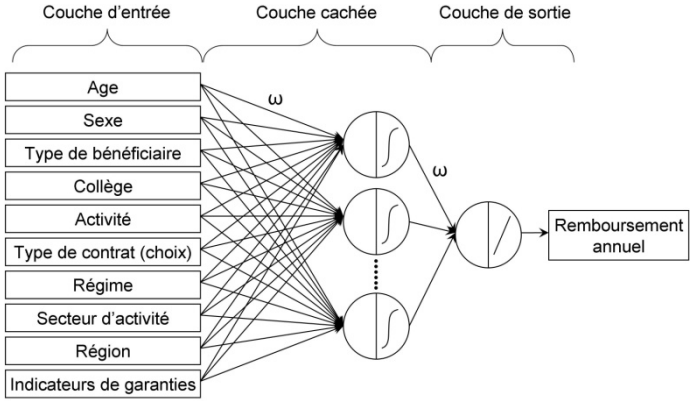



*Figure 5: Schéma du PMC utilisé*

### *3.2.4 Codification des variables catégorielles*

Les variables catégorielles ont dû subir un retraitement afin qu'elles puissent être prises en compte dans le réseau. Chacune de ces variables disposant de $n$ modalités possibles est dupliquée en $n-1$ variables artificielles reprenant ses modalités. Par convention, la dernière modalité (par ordre alphabétique) n'est pas reprise puisqu'elle est directement liée aux autres.

Le mode de conversion retenu s'effectue selon les deux cas de figure suivants :

- si la valeur de la variable à coder est différente de la dernière modalité : la variable qui correspond prend la valeur $1$, les $n-2$ variables artificielles n'indiquant pas la bonne modalité valent 0,
- si la valeur de la variable à coder est celle de la dernière modalité : toutes les variables artificielles valent -1.

Par exemple, la variable « type de bénéficiaire » qui comporte 3 modalités distinctes (adhérent, conjoint ou enfant) a été convertie en 2 variables (bénéficiaire=adhérent, bénéficiaire=conjoint).

La codification est la suivante :

| Type de bénéficiaire | Variables artificielles | |
|---|---|---|
| | bénéficiaire=adhérent | bénéficiaire=conjoint |
| Adhérent | 1 | 0 |
| Conjoint | 0 | 1 |
| Enfant | -1 | -1 |

### *3.2.5 Normalisation des variables*

Les variables continues utilisées n'ont pas forcément des valeurs du même ordre de grandeur. Des écarts importants peuvent nuire aux algorithmes d'apprentissage en les rendant inefficaces. Un prétraitement peut consister à les normaliser en effectuant un changement de variables remplaçant chacune d'entre elles par leur transformation en variable centrée réduite.

Cette solution a été retenue, la transformation pour la variable $x$ de l'exemple $i$, notée $x_i$, est la suivante :

$$x_i \rightarrow \frac{x_i - \overline{x}}{\sigma(x)} \text{ avec } \overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \text{ et } \sigma(x) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

Les variables catégorielles, quant à elles, ne subissent aucune normalisation.

### *3.2.6 Initialisation des poids*

Avant de lancer l'algorithme, les valeurs initiales des poids et des biais ont besoin d'être définies. Si elles ne sont pas correctement fixées, il peut y avoir d'importantes répercussions sur la vitesse d'apprentissage ainsi que sur la capacité de généralisation.

La stratégie retenue s'inspire de celle décrite par Smieja [SMI91] :

- Les poids des connexions de la couche cachée vers celle de sortie sont initialisés à zéro, le biais de la couche de sortie est initialisé à la valeur moyenne de la cible, soit 392,14.
- Pour les poids des connexions de la couche d'entrée vers la couche cachée, un tirage aléatoire est effectué selon une loi normale centrée réduite. Les poids sont ensuite ajustés en les multipliant par $\frac{scale}{d_{in}}$ où $d_{in}$ est le nombre de connexions entrantes dans le neurone (à l'exception des biais) et $scale$ un facteur d'échelle valant $0,1$. Pour appliquer la technique Early Stopping, il est généralement conseillé de prendre un facteur d'échelle de faible valeur comme $0,1$ ou $0,01$.

Pour une architecture donnée, l'initialisation aléatoire des poids est conservée pour toutes les simulations afin qu'elles puissent être réalisées dans les mêmes conditions.

### *3.2.7 Entraînements préliminaires*

Les différents poids et biais finaux du réseau entraîné dépendent de leurs valeurs initiales. Pour réduire le risque de convergence vers un mauvais optimum local, plusieurs entraînements préliminaires peuvent être répétés sur quelques itérations en partant de différentes valeurs initiales tirées aléatoirement. Celui qui offre les meilleurs résultats est retenu. Les poids correspondants servent, à leur tour, pour initialiser la phase d'entraînement. Ainsi, pour éviter les problèmes de convergence, une phase d'entraînement préliminaire de 50 lancés sur 20 itérations est appliquée dans toutes les simulations.

### *3.2.8 L'apprentissage*

La base a été scindée en trois aléatoirement :

- Entraînement (40 %),
- Validation (30 %),
- Test (30 %).

La fonction de coût à minimiser retenue est l'erreur quadratique. La technique

d'apprentissage de base est l'algorithme de rétropropagation du gradient de l'erreur avec un pas de 0,01. Le critère d'arrêt est la technique Early Stopping avec un maximum de 1 000 itérations.

Le processus d'apprentissage est le suivant :

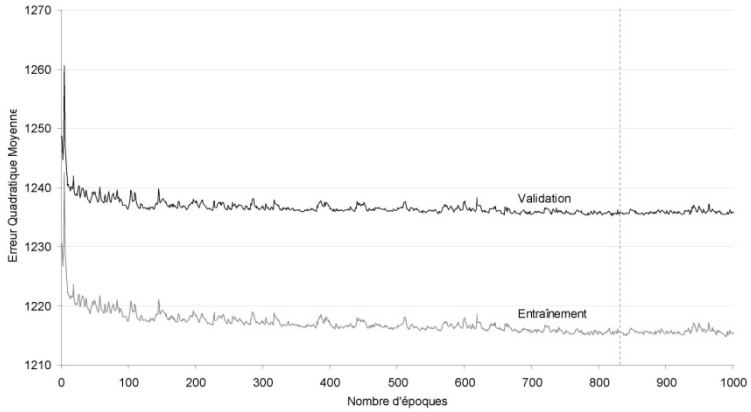



*Figure 6: 1 couche cachée, 20 neurones, Backprop batch, pas de 0.01 et momentum nul*

L'algorithme semble avoir convergé vers un minimum, l'erreur sur la base d'entraînement ne diminue quasiment plus, de même que celle de la base de validation. Le fait que les données soient fortement entachées d'aléa explique l'absence de sur-apprentissage sur la base de validation (l'erreur ne remonte pas clairement après avoir atteint son minimum). Le minimum sur la base de validation est atteint à la 962ème itération, la valeur de l'erreur quadratique moyenne sur la base d'apprentissage est de 1 214,25 et 1 235,27 sur la base de validation. En stoppant l'apprentissage à ce minimum d'erreur sur la base de validation, les coefficients du premier modèle sont définis. En appliquant celui-ci à la base de test, l'erreur obtenue est de 1 286,03. Cette valeur reflète la qualité du modèle lâché dans la nature sur des données inconnues.

Cette première modélisation va servir de référence. Partant de celle-ci, ses différents paramètres vont pouvoir être étudiés un par un afin de les ajuster au mieux selon les résultats obtenus.

## 3.3 Recherche du meilleur modèle

### *3.3.1 Simulations selon plusieurs variantes*

Afin de pouvoir comparer les différentes variantes de modèles le plus objectivement possible, les simulations sont établies à partir des mêmes bases d'entraînement, de validation et de test. La présentation aléatoire des exemples est définie une fois pour toute selon le même ordre lors de chaque apprentissage (même graine de départ).

Les simulations ont été réalisées sous SAS à partir de l'interface *Entreprise Miner* et de la procédure *PROC NEURAL*. Les temps de traitement avec la technique d'apprentissage RPROP sont de l'ordre de 15 heures pour 1 000 itérations et 50 lancés préliminaires de 20 itérations. Ils peuvent être réduits à 4 heures en se limitant à 500 itérations et 5 lancés préliminaires de 20 itérations sans pour autant remettre en cause la pertinence des résultats obtenus. L'utilisation des traitements parallélisés peut être une bonne alternative pour réduire significativement les temps de traitement (disponible sous *SAS* version 9.3).

Les différentes variantes testées sont : le nombre de couches cachées, le nombre de neurones cachés, la présence de liaisons directes, la fonction d'activation, la fonction coût, la technique d'apprentissage, le pas d'apprentissage et le momentum. Le réseaux de neurones ayant obtenu les meilleurs résultats a également été comparé à un réseau de type "RBF" (Radial Basis Functions).

A titre d'exemple, voici le test permettant de choisir la technique d'apprentissage parmi les six proposées dont la synthèse des résultats est la suivante :

| | Méthode | | | | | |
|---|---|---|---|---|---|---|
| | Backprop | Rprop | Quickprop | Gradients Conjugués | Levenberg-Marquardt | Quasi - Newton |
| EQM Min (validation) | 1 234,54 | 1 234,31 | 1 234,79 | 1 234,90 | 1 234,93 | 1 235,51 |
| Epoque | 367ème | 253ème | 979ème | 106ème | 18ème | 31ème |
| EQM (test) | 1 285,69 | 1 285,49 | 1 285,63 | 1 285,59 | 1 285,66 | 1 286,42 |

La méthode « RPROP » est celle qui offre les meilleurs résultats en obtenant la plus faible erreur sur la base de validation. En conséquence, elle sera sélectionnée pour le modèle.

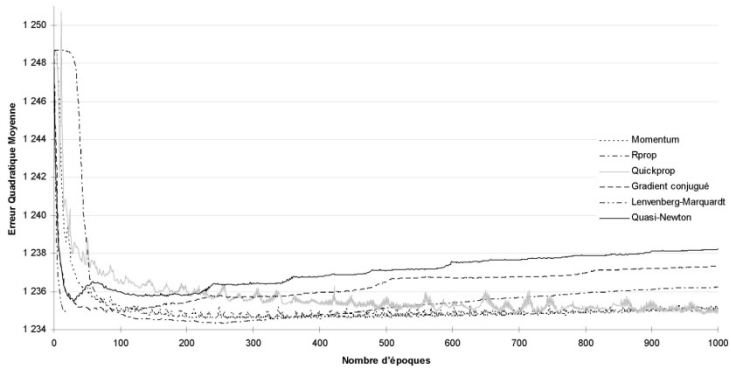



*Figure 7: Validation - RPROP vs Mom. vs Quickp. vs Grad. C. vs Lev.-M. vs Q.-New.*

A l'issue du test des variantes, le modèle sélectionné possède les caractéristiques suivantes :

- Architecture : P.M.C.
- Liaisons directes : non
- Nombre de couches cachées : 1
- Nombre de neurones cachés : 40
- Fonction d'activation : logistique
- Fonction d'erreur : quadratique
- Technique d'apprentissage : RPROP
- Arrêt : 294ème époque
- E.Q.M. min : 1 285,01 en validation (et 1 234,17 en apprentissage)
- Elagage : non

Remarque : Sept variantes de réseaux R.B.F. ont été testées et n'ont pas apporté de meilleurs résultats que les P.M.C.

La synthèse de l'ensemble des variantes testées est la suivante :

| Couches cachées | Neurones cachés | Liaisons directes | Fonction activation | Fonction coût | Technique d'apprentissage | Pas | Moment | EQM (test) |
|---|---|---|---|---|---|---|---|---|
| 1 | 10 | non | tanh | E.Q.M. | Retropropagation | 0,01 | 0 | 1286,72 |
| 1 | 20 | non | tanh | E.Q.M. | Retropropagation | 0,01 | 0 | 1286,03 |
| 1 | 30 | non | tanh | E.Q.M. | Retropropagation | 0,01 | 0 | 1286,1 |
| **1** | **40** | **non** | **tanh** | **E.Q.M.** | **Retropropagation** | **0,01** | **0** | **1285,75** |
| 1 | 50 | non | tanh | E.Q.M. | Retropropagation | 0,01 | 0 | 1285,83 |
| **1** | **40** | **non** | **tanh** | **E.Q.M.** | **Retropropagation** | **0,001** | **0,8** | **1285,69** |
| 2 | 40 | non | tanh | E.Q.M. | Retropropagation | 0,001 | 0,8 | 1286,1 |
| 1 | 40 | non | tanh | E.Q.M. | Retropropagation | 0,1 | 0 | $+\infty$ |
| 1 | 40 | non | tanh | E.Q.M. | Retropropagation | 0,05 | 0 | 1285,86 |
| 1 | 40 | non | tanh | E.Q.M. | Retropropagation | 0,01 | 0 | 1285,75 |
| **1** | **40** | **non** | **tanh** | **E.Q.M.** | **Retropropagation** | **0,001** | **0** | **1285,68** |
| 1 | 40 | non | tanh | E.Q.M. | Retropropagation | 0,0001 | 0 | 1287,74 |
| 1 | 40 | non | tanh | E.Q.M. | Retropropagation | 0,001 | 0 | 1285,68 |
| 1 | 40 | non | tanh | E.Q.M. | Retropropagation | 0,001 | 0,5 | 1285,53 |
| 1 | 40 | non | tanh | E.Q.M. | Retropropagation | 0,001 | 0,7 | 1285,71 |
| **1** | **40** | **non** | **tanh** | **E.Q.M.** | **Retropropagation** | **0,001** | **0,8** | **1285,68** |
| 1 | 40 | non | tanh | E.Q.M. | Retropropagation | 0,001 | 0,9 | 1285,74 |
| 1 | 40 | non | tanh | E.Q.M. | Retropropagation | 0,001 | 0,8 | 1285,68 |
| **1** | **40** | **non** | **tanh** | **E.Q.M.** | **Rprop** | | | **1285,49** |
| 1 | 40 | non | tanh | E.Q.M. | Quickprop | | | 1285,63 |
| 1 | 40 | non | tanh | E.Q.M. | Gradients conjugués | | | 1285,59 |
| 1 | 40 | non | tanh | E.Q.M. | Levenberg-Marquardt | | | 1285,66 |
| 1 | 40 | non | tanh | E.Q.M. | Quasi-Newton | | | 1286,42 |
| **1** | **40** | **non** | **tanh** | **E.Q.M.** | **Rprop ss W.Decay** | | | **1285,49** |
| 1 | 40 | non | tanh | E.Q.M. | Rprop W.Decay | | | 1285,5 |
| 1 | 40 | non | Tan h | E.Q.M. | Rprop | | | 1285,49 |
| 1 | 40 | non | Normale | E.Q.M. | Rprop | | | 1285,32 |
| 1 | 40 | non | Arc Tan | E.Q.M. | Rprop | | | 1285,31 |
| **1** | **40** | **non** | **Logistique** | **E.Q.M.** | **Rprop** | | | **1285,01** |
| **1** | **40** | **non** | **Logistique** | **E.Q.M.** | **Rprop** | | | **1285,01** |
| 1 | 40 | non | Logistique | Huber | Rprop | | | 1291,47 |
| **1** | **40** | **non** | **Logistique** | **E.Q.M.** | **Rprop** | | | **1285,01** |
| 1 | 40 | oui | Logistique | E.Q.M. | Rprop | | | 1285,27 |
| **1** | **40** | **non** | **Logistique** | **E.Q.M.** | **Rprop sans élagage** | | | **1285,01** |
| 1 | 40 | non | Logistique | E.Q.M. | Rprop avec élagage | | | 1285,34 |
| | | | | | **Réseaux R.B.F. (Radial Basis Functions)** | | | |
| 1 | 40 | non | Normale | E.Q.M. | Ordinary RBF-Equal Widths | | | 1298,25 |
| 1 | 40 | non | Normale | E.Q.M. | Ordinary RBF-Unequal Widths | | | 1296,41 |
| 1 | 40 | non | Normale | E.Q.M. | Normalized RBF-Equal Heights | | | 1288,70 |
| 1 | 40 | non | Normale | E.Q.M. | Normalized RBF-Equal Volumes | | | 1289,12 |
| 1 | 40 | non | Normale | E.Q.M. | Normalized RBF-Equal Widths | | | 1287,25 |
| 1 | 40 | non | Normale | E.Q.M. | Normalized RBF-Equal Widths and Heights | | | 1286,10 |
| 1 | 40 | non | Normale | E.Q.M. | Normalized RBF-Unequal Widths and Heights | | | 1291,72 |

### *3.3.2 Incidence du choix de la graine de départ*

100 simulations complètes ont été lancées sur le modèle ayant obtenu les meilleurs résultats indiqué précédemment. Une étude statistique de base permet de s'assurer de la fiabilité de la convergence en testant plusieurs graines de départ.

| Statistique | Apprent. | Validat. | Test |
|---|---|---|---|
| Moyenne | 1 212,87 | 1 234,43 | 1 285,31 |
| Ecart-type | 0,54 | 0,12 | 0,10 |
| Médiane | 1 212,90 | 1 234,43 | 1 285,32 |
| Min | 1 211,68 | 1 234,07 | 1 285,01 |
| Max | 1 214,23 | 1 234,74 | 1 285,54 |

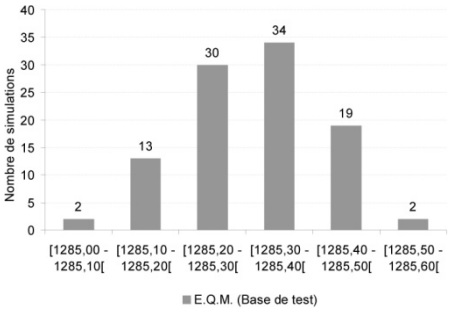



*Figure 8 : Dirtribution des E.Q.M. sur la base de test*

Les résultats des 100 simulations effectuées sont assez proches avec des valeurs d'erreurs comprises entre 1285,01 et 1285,54. Aucune convergence vers un mauvais minimum local n'a pu être constatée.

## 3.4 Analogie avec des modèles classiques de type G.L.M.

En conservant les notations précédentes avec $N_c$ le nombre de neurones cachés, $o_{N_c+1}$ la sortie unique du réseau et $x$ le vecteur des variables en entrée de dimension $n+1$, le modèle retenu a pour expression :

$$o_{N_c+1}(x,w) = \sum_{i \in pred(N_c+1)} \left[ w_{N_c+1,i} . \phi \left( \sum_{j \in pred(i)} w_{ij} x_j + w_{i0} \right) \right] + w_{N_c+1,0} \qquad (1)$$

$$= w_2 . f(W_1 x)$$

$w_2$ est le vecteur des paramètres de la deuxième couche de connexions de dimension $N_C+1$, $W_1$ est la matrice des connexions de la première couche de dimension $N_C+1, n+1$ et $f(.)$ est le vecteur comprenant le biais ainsi que les fonctions effectuées par les neurones cachées de dimension $N_C+1$.

La forme de la deuxième équation fait apparaître une extension naturelle d'un modèle linéaire statique classique : une combinaison linéaire de fonctions non linéaires paramétrées des variables.

# 4. COMPARATIF AVEC D'AUTRES MODELES ET VALIDATION DES RESULTATS

## 4.1 Comparatif avec des modèles avancés

Le réseau de neurones sélectionné précédemment a été confronté à deux autres modèles avancés pour s'assurer du haut niveau de ses performances : un Modèle Linéaire Généralisé (G.L.M.) et un modèle Memory Base Reasoning (M.B.R.).

### *4.1.1 Comparatif avec un Modèle Linéaire Généralisé (G.L.M.)*

Dans le modèle étudié, la variable à expliquer est supposée suivre une loi normale avec comme une fonction de lien la fonction identité $f(y) = y$. Compte tenu du grand nombre de variables, une infinité de modèles linéaires peuvent être envisagés. Les différentes interactions entre les variables ainsi que les effets polynomiaux engendrent une multitude de combinaisons possibles. Pour pallier ce problème, trois méthodes de sélection automatique de variables existent : la méthode d'élimination (Backward), la méthode de sélection (Forward) et la méthode mixte (Stepwise). Elles consistent à ajouter ou à retirer des effets jusqu'à ce qu'aucune variable ne puisse entrer ou sortir du modèle ou bien qu'un critère d'arrêt soit satisfait. Ne pouvant tester l'intégralité des modèles, une présélection de variables candidates a dû être réalisée.

Les résultats ainsi obtenus sont les suivants :

| | Type de modèle | | | | |
|---|---|---|---|---|---|
| | Facteurs simples | + Interac. d'ordre 2 | Méthode Backward | Méthode Forward | Méthode Stepwise |
| EQM Min (validation) | 1 237,18 | 1 237,33 | 1 235,91 | 1 235,88 | 1 235,88 |
| EQM (test) | 1 288,09 | 1 288,30 | 1 286,64 | 1 286,82 | 1 286,82 |

Les résultats de ce modèle n'ont pas permis de battre les performances de la modélisation neuronale.

Sur la base de test, la valeur de l'Erreur Quadratique Moyenne est de 1 286,64 sur le modèle G.L.M. avec la méthode Backward contre 1 285,01 sur le réseau de neurones.

### *4.1.2 Comparatif avec un Modèle Memory Based Reasoning (M.B.R.)*

Cette méthode est appelée aussi « $k$ plus proches voisins » (en anglais, K.N.N. $K$ - Nearest Neighbor) inventée en 1986 par Stanfill et Waltz [STA86] puis reprise par Weiss et Kulikowski en 1991 [WEI91]. L'idée de base est d'utiliser les exemples déjà connus pour

effectuer une prédiction de manière simple sur des nouveaux exemples. Cette technique est adaptée pour des problèmes de classification ou de régression d'une fonction inconnue. La spécificité de cette méthode est qu'aucun modèle n'est induit à partir des données. L'intégralité des exemples est stockée en mémoire en l'état. Pour prédire à partir d'un nouvel exemple, l'algorithme sélectionne ses $k$ plus proches voisins dans la base mémorisée, la réponse affectée est la plus fréquente dans le cadre d'une classification ou la valeur moyenne de la fonction à prédire. Les deux seuls paramètres sont le nombre des voisins les plus proches $k$ et la fonction de similarité qui permet de définir un critère de proximité entre les exemples $d(x_i, x_j) = \sqrt{\sum_{r=1}^{n}\left(a_r(x_i) - a_r(x_j)\right)^2}$ .

Cette technique est testée avec la distance euclidienne et k=16 (nombre de voisins les plus proches) et la méthode de recherche des plus proches voisins Rd-Tree (Reduced Dimentionality Tree).

Les Erreurs Quadratiques Moyennes obtenues sur les bases d'entraînement, de validation et de test ont respectivement les valeurs suivantes : 1 263,41, 1 319,98 et 1 363,90.

| | Modélisation | | |
|---|---|---|---|
| | Réseau de neurones | G.L.M. | Memory Based Reasoning |
| EQM Min (validation) | 1 234,17 | 1 235,91 | 1 319,98 |
| EQM (test) | 1 285,01 | 1 286,64 | 1 363,90 |

Le modèle de réseau de neurone reste celui qui offre les meilleures performances. Sur la base de test, la valeur de l'Erreur Quadratique Moyenne reste toujours plus faible sur le modèle de réseau de neurones avec 1285,01. Le réseau de neurones n'est pas battu par les deux modèles G.L.M. et Memory Based Reasoning, démontrant ainsi son utilité.

L'exemple ci-dessous est un comparatif entre les prédictions des trois modèles dans le cadre d'un adhérent actif, de sexe masculin, vivant en Ile-de-France, travaillant dans le secteur de services aux entreprises, bénéficiant d'un contrat ensemble du personnel milieu de gamme :

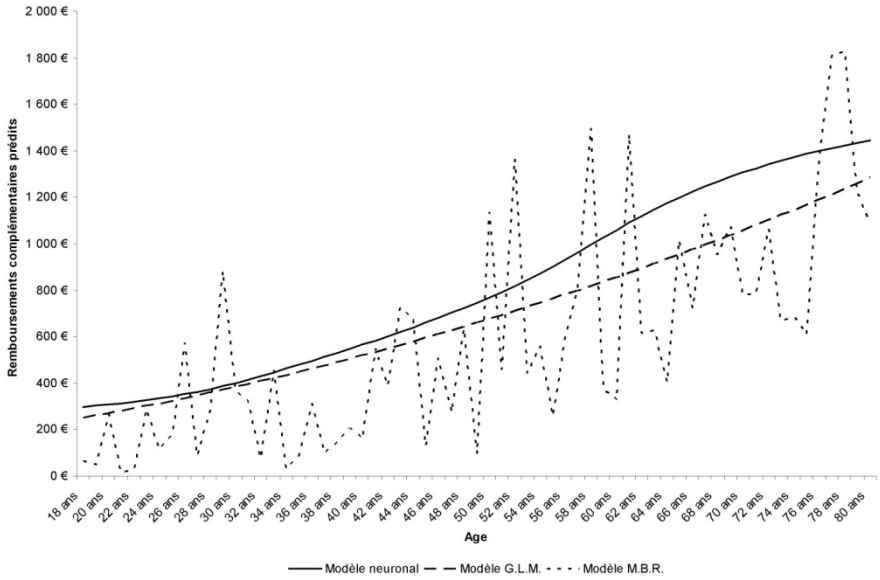



*Figure 9: Prédictions selon les trois modèles en fonction de l'âge*

### 4.2 Tests sur de nouvelles données

Les trois modèles ont été testés dans des conditions réelles d'utilisations. Après avoir été bâtis sur une base de 2007, ils sont testés avec la consommation médicale de 2009 issue de 13 nouveaux contrats entrés en portefeuille après 2007. Une hypothèse de dérive de la consommation médicale à 1,83% par an, correspondant à l'évolution observée entre les exercices 2006 et 2007, a été anticipée.

Les primes pures obtenues en moyenne par entreprise sont les suivantes :

| | Tarif annuel moyen par famille | | | |
|---|---|---|---|---|
| | Réel | R.N.A. | G.L.M | M.B.R. |
| Moyenne | 1 028 | 1 069 | 984 | 1 122 |
| Ecart/Réel | | 3,98% | -4,27% | 9,06% |
| Ecart-type | | 148 | 257 | 351 |

La prédiction du modèle neuronale est la plus proche de la réalité avec un écart de 4% au global. Avec la plus faible valeur d'écart-type, le modèle neuronale différencie le mieux les tarifs des entreprises testées.

### 4.3 Tests de cohérence

Pour s'assurer du bon comportement du modèle, la cohérence des prédictions a été testée en faisant varier un jeu de paramètres sur une population de référence : bénéficiaires vivant en Ile-de-France dont l'adhérent travaille dans le secteur de services aux entreprises et un contrat ensemble du personnel ayant de bonnes garanties. A titre d'exemple, voici le test des critères principaux suivants : le sexe, l'âge et le type de bénéficiaire.

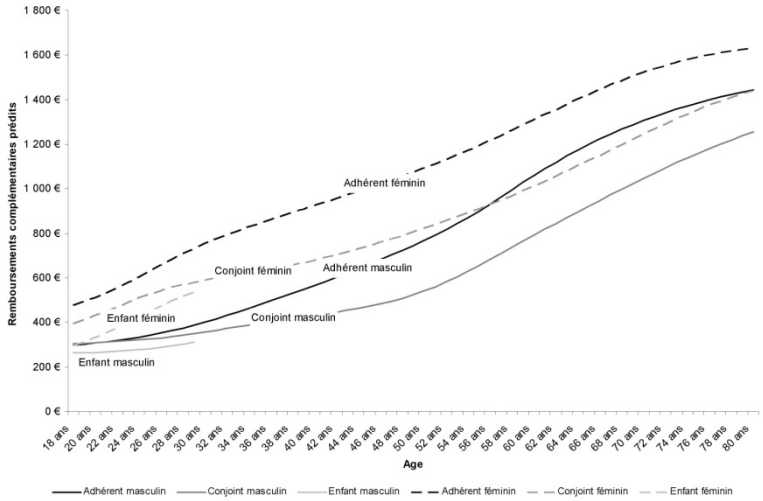



*Figure 10: Effets âge, sexe et type de bénéficiaire*

Ce premier test est concluant car le modèle reproduit bien les tendances attendues :

- la consommation croît avec l'âge,
- les femmes consomment davantage que les hommes,
- cet écart diminue pour les âges élevés,
- les adhérents consomment plus que les conjoints.

### 4.4 Comparatif avec une méthode classique

#### *4.4.1 Modèles de type fréquence × coût moyen*

Les techniques classiques de tarification santé s'inspirent souvent de la méthode élaborée en 1968 par le B.C.A.C.[1], actualisée en 1981, puis en 2001. Elles reposent sur le principe de décomposition du risque entre la fréquence et le coût moyen.

[1]Bureau Commun d'Assurances Collectives

Ce modèle a le mérite d'être assez simple d'utilisation et d'être employé par une majorité d'entreprises d'assurance et de bénéficier d'un coût par poste. Implicitement dans cette approche, les dépendances entre les fréquences, les frais réels et les garanties ne sont pas supposées exister. Or dans les faits, ces grandeurs sont étroitement liées. En conséquence, la précision des résultats obtenus pourrait être sensiblement améliorée.

### *4.4.2 Protocole de comparaison*

Afin de vérifier en pratique l'intérêt du modèle, il a été testé dans des conditions réelles d'utilisation. Ces résultats ont été comparés aux tarifs pratiqués obtenus par des méthodes traditionnelles.

La prévision de la consommation médicale sera appliquée sur la population active des 13 plus grandes entreprises rentrées en portefeuille à partir de 2008 (pour exclure celles ayant servi à bâtir le modèle). Ce panel comprend 10 secteurs d'activité, 24 contrats, près de 24 000 adhérents et 59 000 bénéficiaires.

L'information de chaque bénéficiaire est reprise sur une ligne par individu. Elle comprend le niveau de garantie de chaque poste ainsi que les caractéristiques démographiques. Ces lignes sont ensuite présentées au modèle, une par une, pour être tarifées.

La démarche consiste à comparer les trois éléments suivants :

- le montant de la prime pure réellement encaissée = tarif appliqué ( $\Pi_{applique'}$ )
- le montant de prestations prédit par le modèle = tarif estimé ( $\Pi_{pre'dit}$ )
- le montant réel des prestations versées = tarif d'équilibre ( $\Pi_{e'quilibre}$ )

Le passage de la prime commerciale à la prime pure réellement encaissée se fait en retirant la C.M.U. [1] en l'assimilant à une taxe (5,90% en 2009) et les chargements :

$$\Pi_{applique'} = \Pi_{brute\ commerciale} \times \frac{1 - chargements}{1 + CMU}$$

Les prestations versées de survenance 2009 sont arrêtées à fin juin 2010. A cette date d'observation un taux de 1,70% de P.S.A.P.[2] a été appliqué pour reconstituer un exercice complet en intégrant la queue de sinistres :

$$\Pi_{e'quilibre} = Prestations\ 2009_{30-06-10} \times (1 + PSAP_{30-06-10})$$

Le tarif prédit par le modèle a été majoré d'une hypothèse de dérive de la

[1] Couverture Maladie Universelle

[2] Provisions pour Sinistres A Payer

consommation médicale sur deux années. Elle correspond à l'évolution moyenne des remboursements observée sur le portefeuille entre 2006 et 2007 qui était de 1,83% par an :

$$\Pi_{pre'dit} = \Pi_{mod'ele\ RN} \times (1 + de'rive_{2006-2007})^2$$

Les montants des primes pures obtenus sur ces 13 entreprises sont les suivants :

| Entreprise | Effectifs | Montants totaux de primes | | | Tarifs annuels par famille | | |
|---|---|---|---|---|---|---|---|
| | | $\Pi_{pre'dit}$ | $\Pi_{applique'}$ | $\Pi_{e'quilibre}$ | $\Pi_{pre'dit}$ | $\Pi_{applique'}$ | $\Pi_{e'quilibre}$ |
| Entrep. A | 7 748 | 9 309 383 | 7 474 523 | 8 845 661 | 1 202 | 965 | 1 142 |
| Entrep. B | 3 257 | 3 347 102 | 2 366 063 | 3 810 500 | 1 028 | 726 | 1 170 |
| Entrep. C | 2 963 | 2 539 634 | 2 231 838 | 2 563 629 | 857 | 753 | 865 |
| Entrep. D | 2 176 | 2 657 857 | 2 157 608 | 2 161 029 | 1 221 | 992 | 993 |
| Entrep. E | 1 852 | 2 363 759 | 2 207 763 | 2 388 484 | 1 276 | 1 192 | 1 290 |
| Entrep. F | 1 391 | 1 253 882 | 1 106 076 | 994 590 | 901 | 795 | 715 |
| Entrep. G | 981 | 1 094 303 | 651 272 | 883 419 | 1 115 | 664 | 901 |
| Entrep. H | 727 | 612 581 | 480 405 | 571 602 | 843 | 661 | 786 |
| Entrep. I | 726 | 656 352 | 598 981 | 784 584 | 904 | 825 | 1 081 |
| Entrep. J | 540 | 313 301 | 402 180 | 501 080 | 580 | 745 | 928 |
| Entrep. K | 519 | 576 854 | 292 823 | 328 788 | 1 113 | 565 | 634 |
| Entrep. L | 503 | 459 337 | 340 291 | 406 907 | 913 | 677 | 809 |
| Entrep. M | 498 | 353 409 | 326 029 | 320 747 | 710 | 655 | 644 |
| Tot./moy. | 23 881 | 25 537 761 | 20 635 858 | 24 561 027 | 1 069 | 864 | 1 028 |

Sur l'ensemble des effectifs couverts, dans 79% des cas, le modèle de réseau de neurones fournit de meilleures estimations du risque. Au global, le tarif prédit est supérieur de 4% au tarif d'équilibre alors que les tarifs pratiqués sont inférieurs de 16%. L'application du modèle de tarification neuronale dans des conditions normales d'utilisation a été concluante en apportant des résultats plus précis que ceux appliqués.

## 5. CONCLUSIONS

Une méthodologie complète utilisant la technique des réseaux de neurones a pu être expérimentée pour tarifer des contrats collectifs frais de santé. Cette nouvelle approche permet de tirer profit de la souplesse et de la précision que leur confère leur propriété fondamentale d'approximateurs universels parcimonieux.

La confrontation de la démarche neuronale proposée avec d'autres outils mathématiques avancés s'est avérée être concluante. Les Modèles Linéaires Generalisés et Memory Based Reasoning ne sont pas parvenus à battre ses performances, ce qui a conforté le choix porté pour cette technique.

Elle a pu mettre en avant sa valeur ajoutée en étant appliquée dans des conditions réelles d'utilisation sur un panel des 13 entreprises représentant près de 60 000 bénéficiaires. Dans 79% des cas, les résultats obtenus par le modèle de réseau de neurones ont été plus proches de la réalité que les tarifs appliqués provenant d'outils classiques de tarification. Le coût du risque global prédit pour l'ensemble du périmètre ressort supérieur de 4% au tarif d'équilibre alors que les tarifs pratiqués étaient inférieurs de 16%.

Cette nouvelle approche tarifaire a pu démontrer l'intérêt d'une modélisation neuronale car celle-ci offre un niveau de précision supplémentaire comparativement aux méthodes traditionnellement utilisées.

La modélisation neuronale présente un avantage majeur dans le pilotage de régimes frais de santé. En offrant une vision plus précise du risque, elle contribue à pérenniser dans le temps sa gestion. Elle permet de répondre à la nécessité pour l'entreprise de maîtriser ses budgets de protection sociale dans un environnement très concurrentiel aggravé par la crise économique et un contexte social parfois difficile.

En synthèse :

| ***Caractéristiques*** | **Méthode proposée** | **Méthode classique** |
|---|---|---|
| Modélisation | Réseau de Neurones | Fréquence × Coût moyen |
| Approche | Tarif global | Tarif par poste |
| Influences des garanties sur les frais réels et les fréquences | Prises en compte | Non prises en compte |
| Liaisons entre les garanties | Prises en compte | Non prises en compte |
| Pouvoir explicatif des coefficients | Non | Oui |
| Précision des prédictions | Haute | Moyenne |

**REFERENCES**

[AOU10] AOUIZERATE J.M. (2010) : *Alternative neuronale en tarification santé*, Mémoire d'actuariat, C.N.A.M. (Conservatoire National des Arts et Métiers)

[BAR93] BARRON A. (1993) : Universal Approximation Bounds for Superposition of a Sigmoidal Function, *IEEE Transactions on Information Theory Society*, vol. 39, p. 930-945.

[BOR07] BORNE P., HAGGEGE J., BENREJEB M. (2007) : *Les réseaux de neurones : présentation et applications*, Méthodes et pratiques de l'ingénieur, Editions Technip, Paris

[BOU07] BOUKHETALA K., YAHIAOUI M., LAADJEL T. (2007) : Une approche de tarification en assurance automobile par réseaux de neurones, *Revue des Nouvelles Technologies de l'Information, Data Mining et Apprentissage Statistique : applications en assurance, banque et marketing*, Cepadues-Editions, p. 189-203

[CUL43] MAC CULLOCH W.S., PITTS W.H. (1943) : A Logical Calculus of the Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biophysics*, The University of Chicago Press, vol. 5, p.115-133

[DRE08] DREYFUS G., MARTINEZ J.M., SAMUELIDES M., MIRTA B. (2008) : *Apprentissage statistique*, Editions Eyrolles

[DRE10] DREYFUS G. (2010) : *Neural Networks: Methodology and Applications*, Springer

[DUG03] DUGAS C., BENGIO Y., CHAPADOS N., VINCENT P., DENONCOURT G., FOURNIER C. (2003) : Statistical Learning Algorithms Applied to Automobile Insurance Ratemaking, *Intelligent Techniques for The Insurance Industry: Theory and Applications*, eds. L. Jain and A.F. Shapiro, World Scientific

[FAU94] FAUSETT LAURENE V. (1994) : *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*, Prentice-Hall

[HAS08] HASTIE T., TIBSHIRANI R., FRIEDMAN J. (2008) : *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second edition, Springer, p. 389-415

[HEB49] HEBB D.O. (1949) : *The Organization of Behavior: A Neuropsychological Theory*, Wiley, New-York

[HOP82] HOPFIELD J.J. (1982) : Neural Networks and Physical Systems with Emergent Collective Computational Abilities, *Proceedings of the National Academy of Sciences*, vol. 79, No. 8., p. 2554-2558

[HOR93] HORNIK K. (1993) : Some New Results on Neural Network Approximation, *Neural Networks*, Elsevier Science Ltd., vol. 6, no 8, p. 1069-1072

[HSU08] HSU S., LIN C., YANG Y. (2008) : Integrating Neural Networks for Risk-Adjustment Models, *Journal of Risk and Insurance*, American Risk and Insurance Association inc., vol. 75, Issue 3, p. 617-642, September 2008.

[KIT02] KITCHENS F.L., BOOKER Q.E., REBMAN C.M. Jr. (2002) : An Application of Neural Networks to Insurance Underwriting, *33rd Annual Conference of the Decision Sciences Institute*, Southwest Region, St. Louis Missouri

[KIT09] KITCHENS F.L. (2009) : Financial Implications of Artificial Neural Networks in Insurance Underwriting, *International Journal of Electronic Finance*, Inderscience Publishers, vol. 3, N°3, p. 311 -319

[LET59] LETTVIN J.Y., MATURANA H.R., MAC CULLOCH W.S., PITTS W.H. (1959) : What the Frog's Eye Tells the Frog's Brain, *Proceedings of the Institute of Radic Engineers*, vol. 47, N°11, p. 1940-1951

[LIU11] LIU D., ZHANG H., POLUCARPOU M., ALIPPI C. and HAIBO H. (2011) : Advances in Neural Networks, *8th International Symposium on Neural Networks*, ISNN 2011, Guilin, China, Springer, Proceedings, Part III, May 29-June 1

[LIO08] LIOU F.M., TANG Y.C., CHEN J.Y. (2008) : Detecting Hospital Fraud and Claim Abuse through Diabetic Outpatient Services, *Health Care Management Science*, Springer, vol. 11, p. 353-358

[LIN08] LIN Y.J., HUANG C.S., LIN C.C. (2008) : Determination of Insurance Policy Using Neural Networks and Simplified Models with Factor Analysis Technique, *WSEAS Transactions on Information Science and Applications*, vol. 5, Issue 10, October 2008

[LOW96] LOWE J., PRYOR L. (1996) : Neural Networks v. GLMs in Pricing General Insurance, *General Insurance Convention 1996*, vol. 2, p. 417-438

[MAT05] MATIGNON R. (2005) : *Neural Network Modeling Using SAS Enterprise Miner*, SAS® Institute Inc., Authorhouse

[MIN69] MINSKY M., PAPERT S. (1969) : Perceptrons Cannot Learn Arbitrary Rules, *Perceptrons* , Cambridge MA: MIT Press, Introduction, p. 1-20

[MUL06] MULQUINEY P. (2006) : Artificial Neural Networks in Insurance Loss Reserving, *JCIS*, DBLP, Atlantis Press 2006

[ORT06] ORTEGA P.A., FIGUEROA C.J., RUZ G.A. (2006) : A Medical Claim Fraud/Abuse Detection System based on Data Mining: A Case Study in Chile, *DMIN'06, The 2006 International Conference on Data Mining*, p. 224-231

[PAG11] PAGLIA A. PHELIPPE-GUINVARC'H M.V. (2011) : Tarification des risques en assurance non-vie, une approche par modèle d'apprentissage statistique, *Bulletin Français d'Actuariat*, Institut des Actuaires, Vol. 11, N°22, juillet-décembre 2011, p. 49-81

[PEL98] PELESSONI R., PICECH L. (1998) : Some applications of unsupervised neural networks in rate making procedure, *General Insurance Convention & ASTIN Colloquium 1998*, Glasgow, vol. 2, Faculty and Institute of Actuaries, p. 550-567

[PEL99] PELESSONI R., PICECH L. (1999) : An application of Unsupervised Neural Networks in General Insurance: the Determination of Tariff Classes, *Proceedings of NEU99*, Venezia, Dip. di Matematica Appl. Università Ca'Foscari di Venezia, p. 64-75

[REN06] RENNARD J.P. (2006) : *Réseaux neuronaux : une introduction accompagnée d'un modèle Java*, Vuibert

[ROS57] ROSENBLATT F. (1957) : The Perceptron: a Perceiving and Recognizing Automaton, Project PARA. Technical Report N° 85-460-1, Cornell Aeronautical Laboratory, Jan. 1957

[ROS58] ROSENBLATT F. (1958) : The Perceptron: a Probabilistic Model for Information Storage and Organization in the Brain, *Psychological Review*, vol. 65, p. 386-408

[RUM86] RUMELHART D.E, HINTON G.E., WILLIAMS R.J. (1986) : Learning Internal Representations by Error Propagation, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, M.I.T. Press, vol. 1, p. 318-362.

[SIM09] SIMON S., HAYKIN (2009) : *Neural Networks and Learning Machines*, vol. 10, Prentice Hall

[SMI91] SMIEJA F.H. (1991) : Hyperplane "Spin" Dynamics, Network Plasticity and Back-Propagation Learning, GMD Report, GMD, St Augustin, Germany

[SMI02] SMITH. K., GUPTA J., JOHNSON J. (2002) : Predicting Automobile Insurance Losses Using Artificial Neural Networks, *Neural Networks in business: Techniques and Applications*, Idea Group Publishing

[STA86] STANFILL C., WALTZ D. (1986) : Toward Memory-Based Reasoning, *Communications of the ACM*, Vol. 29, N°12, p. 1213-1228

[SUH12] SUH S.C. (2011) : *Practical Applications of Data Mining*, Jones and Bartlett Learning, p. 226-278

[TAN10] TANG H., TAN K.C., YI Z. (2010) : *Neural Networks: Computational Models and Applications*, Springer

[TUF10] TUFFERY S. (2010) : *Data mining et statistique décisionnelle: L'intelligence des données*, Editions Technip, troisième édition, p. 209-223

[WEI91] WEISS S.M., KULIKOWSKI C. (1991) : *Computer Systems That Learn*, Morgan Kaufmann

[WER74] WERBOS P.J. (1974) : *Beyond Regression: New Tools for Prediction and Analysis in the Behavorial Sciences*, Doctoral dissertation, Applied mathematics, Harvard university, Boston

[WID60] WIDROW B., HOFF M.E. (1960) : Adaptive Switching Circuits, *1960 IRE WESCON Convention Record*, vol. 4, p. 96-104, Los Angeles

[WIT11] WITTEN I.H., FRANK E., HALL M.A. (2011) : *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufman, p. 129, 232-242, 469-472

[YEH10] YEH C.P., CHANG W.H., CHIANG S.M., HSU C.W. (2010) : Forecasting the Expenditures of Taiwan National Health Insurance: Back-Propagation Neural Network and ARIMA Models Analysis, *International Conference on Business And Information (BAI 2010)*, Kitakyushu, July 5-7

[YEO05] YEO A.C. (2005) : Neural Networks for Automobile Insurance Pricing, *Encyclopedia of Information Science and Technology*, Mehdi Khosrow-Pour, IGI Global, p. 2095-2099