

Une approche de tarification en assurance automobile par Réseaux de Neurones

K. Boukhetala, M. Yahiaoui et T.Laadjel

Département de Probabilités-Statistique et Recherche Opérationnelle
Bp 32, El Alia, Bab-Ezzouar, USTHB, Alger.
kboukhetala@usthb.dz

Résumé. Une approche d'estimation de la prime de risque, en assurance automobile, en utilisant la technique des réseaux de neurones, est proposée. L'entrée du réseau est constituée d'un vecteur de facteurs de risque significatifs. La sortie est un vecteur de classes de risque appropriées. La règle d'apprentissage que nous proposons, et qui est adaptée aux caractéristiques du problème traité, permet l'affectation d'un client (assuré) qui rentre dans le système, à une classe de risque correspondante. Un intervalle de confiance pour la prime de base est déterminé pour chaque classe de risque.

1 Introduction

En économie moderne, l'assurance joue un rôle primordial dans la conception de stratégies de développement d'un pays. Une telle activité est extrêmement liée à la notion de risque et sa couverture. Les risques assurables nécessitent l'utilisation d'outils scientifiques pour déterminer les règles les plus appropriées à sa gestion. Les sciences actuarielles et le calcul stochastique constituent l'outil indispensable pour l'élaboration des systèmes de tarification rationnels. Le développement de l'outil informatique, ces dernières années, a beaucoup contribué à l'innovation et au développement du métier de l'assurance. Récemment, le problème de modélisation des risques a fait l'objet d'importants travaux de recherche; notamment ceux de (Partrat, C., et Besson, J-L, (2005)), (Rolski, T., et al. (1998)), (Haberman, S., et (Pitacco, E., (1999)), (Necir, A., et Boukhetala, K., (2004))). La détermination d'une bonne prime de risque est le souci de l'actuaire et l'économiste. Le risque d'événement rare présente des difficultés particulières dans sa gestion et nécessite des approches de modélisation appropriées. (Wong, S., (1996)) propose un principe de prime pour ce type de risque, utile pour le calcul de la prime de réassurance. Récemment, un estimateur, asymptotiquement normal, de la prime basée sur le principe de Wong, a été développé par (Necir, A., et Boukhetala, K., (2004)). Une étude sur des données réelles de la sinistralité automobile, au sud de l'Algérie, a été réalisée en utilisant cet estimateur. Une comparaison des risques permet une partition, en classes de risque, de la population des assurées. En Algérie, le système d'assurance est très ancien. La partition de la population d'assurés, en classes de risque, nécessite aujourd'hui un réajustement et une mise à jour, vu que l'environnement automobile a considérablement changé; extension et diversification du parc, multiplication des facteurs d'aggravation du risque, etc... Après réduction,

par regroupement, du nombre de classes de risque du système de tarification actuel de l’automobile en Algérie, nous proposons, dans cet article, une approche scientifique adaptée, qui permet d’affecter un client qui rentre dans le système à une nouvelle classe de risque, en utilisant une approche par les réseaux de neurone. Lorsque le client est affecté, par une certaine règle d’apprentissage du réseau, à une classe correspondante, une prime de risque pourra lui être calculée. L’article est organisé comme suit : La première partie est consacrée à la présentation d’une nouvelle partition, en classes de risque. La seconde partie présente une description de la méthode des réseaux de neurones et son adaptation au problème traité. La dernière partie traite une application de cette approche pour l’estimation, par intervalle de confiance, de la prime de risque. L’approche peut constituer un outil d’aide à la conception d’une stratégie de tarification.

2 Facteurs et classes de risque

Les facteurs de risque les plus significants, retenus par le système étudié, sont le genre, la zone de circulation, l’usage et la puissance du véhicule. D’autres facteurs comme l’âge du permis de conduire, l’âge du véhicule sont considérés comme facteurs aggravants du risque. Pour chacun de ces facteurs de risque, il est associé un ensemble de codes. Un ensemble de classes de risque est considéré par la combinaison de ces codes; comme par exemple, 00103, désigne la classe des véhicules sans remorque, circulant dans la zone nord, à usage affaire de puissance 7 à 10 chevaux. Le système en question comprend 47 classes de risque. Une étude statistique sur des données réelles a montré que cette classification nécessite une mise à jour par un réajustement de l’ensemble de classes de risque. Une classification hiérarchique a donnée que le nombre de classes significantes se réduit à 19. Dans cette nouvelle classification, il est tenu compte de la proximité des classes, afin de conserver certaines hypothèses pratiques liées à l’ancienne partition. Le tableau suivant résume les regroupements obtenus pour cette nouvelle classification proposée.

Nouvelles classes	Agrégat d’anciennes classes
1	00100, 00101
2	00102
3	00103
4	00104, 00105, 00106
5	00110, 00111
6	00112
7	00113
8	00114, 00115, 00116
9	00120, 00121, 00122
10	00123
11	00124, 00125, 00126
12	00130, 00131, 00132
13	00133, 00134, 00135, 00136
14	02* * *, 03* * *, 05107, 06108, 08108
15	04150, 04153, 04154
16	30140, 31140, 32150, 33150
17	34* * *
18	35170, 36170, 38170
19	45170, 46170

TAB. 1 – – agrégation des classes

En se basant sur cette nouvelle classification, qui est tout à fait économique en matière de gestion du système, nous allons proposer une nouvelle approche de tarification en utilisant les réseaux de neurones.

3 Les réseaux de neurones (voir Haykin, S., (1994))

Les réseaux de neurones sont des structures (la plupart du temps simulées par des algorithmes exécutés sur des ordinateurs d'usage général, parfois sur des machines ou même des circuits spécialisés) qui prennent leur inspiration (souvent de façon assez lointaine) dans le fonctionnement élémentaire des systèmes nerveux. Ils sont utilisés essentiellement à résoudre des problèmes de classification, de reconnaissance de formes, d'association, d'extraction de caractéristiques, d'identification, etc. Ils deviennent des compléments aux méthodes classiques, et sont même susceptibles de se substituer à celle-ci avec un taux de succès supérieur.

3.1 Connexionnisme

Les observations dans le monde réel (ce qui est reçu par un système perceptif) sont formées de très nombreuses composantes, de nature variée, ayant des importances en général difficiles à évaluer a priori. En d'autres termes, un stimulus peut être représenté comme un vecteur appartenant à un espace de dimensionnalité élevée, ayant des composantes souvent hétérogènes, et entaché parfois de composantes non significatives. Le rôle d'un système de classification est de traiter l'ensemble de ces composantes, pour en dégager automatiquement les traits significatifs (les caractéristiques), et permettre en dernière étape d'aboutir à une décision, c'est à dire de décider à quelle catégorie ou classe appartient le stimulus initial. Dans l'espace de décision, la sortie est un vecteur de dimensionnalité beaucoup plus faible (souvent même de dimension unité, dans le cas de la décision élémentaire oui/non). Le rôle du système de classification est donc d'opérer une transformation associant un espace d'entrée de forte dimensionnalité où règne un certain indéterminisme, à un espace de sortie de faible dimensionnalité, où l'on désire obtenir des réponses aussi déterministe que possible. La fonction qui fait passer d'un espace à l'autre est très compliquée ou trop difficile à connaître explicitement pour être représentée sous forme mathématique, donc par programmation. Elle n'est accessible que par apprentissage, c'est-à-dire par exemple, par association et comparaison d'un certain nombre d'exemples connus à l'avance (ou connus au fur et à mesure du processus) avec les réponses correctes. La présentation d'exemples permet de modifier progressivement les caractéristiques du système jusqu'à ce qu'un comportement global soit estimé suffisant. Le succès de l'opération n'est évidemment garanti, et l'on peut comprendre intuitivement qu'une quantité insuffisante d'exemples, c'est-à-dire ne contenant pas assez d'information, ne permette pas d'atteindre une configuration satisfaisante.

3.2 Un exemple fondamental : le perceptron

Le perceptron est le premier des réseaux de neurones (cf. figure1). Il fut mis au point par (Rosenblatt, R., (1958)). Le but du perceptron est d'associer des configurations (des formes) en entrée à des réponses. Le perceptron se compose de deux couches : la rétine et la couche de sortie qui donne la réponse correspondante à la stimulation présente en entrée. Les cellules

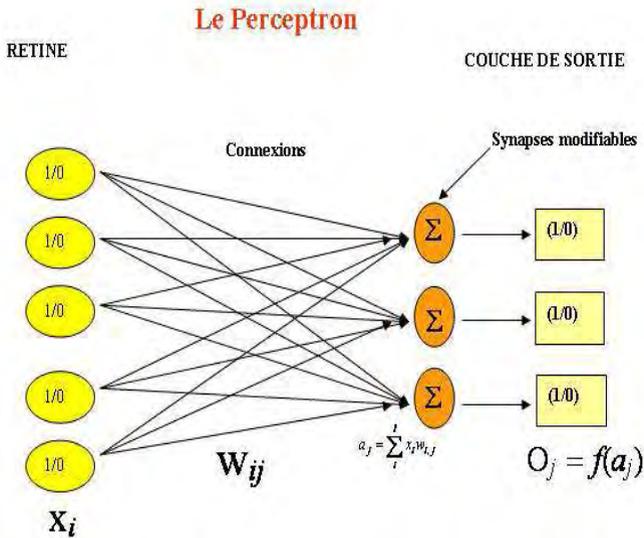


FIG. 1 -- *Le perceptron*

de la première couche répondent en oui/non. La réponse 'oui' correspond à une valeur '1' et la réponse 'non' correspond à une valeur '0' à la sortie du neurone. Les cellules d'entrée sont reliées aux cellules de sortie grâce à des synapses d'intensité variable. L'apprentissage du perceptron s'effectue en modifiant l'intensité de ces synapses.

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

$$a_j = \sum_i x_i w_{ij}$$

où

- a_j : activation totale de la $j^{\text{ème}}$ cellule
- x_i : valeur (0 ou 1) de la $i^{\text{ème}}$ cellule de la rétine
- $w_{i,j}$: intensité de la connexion entre la $i^{\text{ème}}$ cellule d'entrée et la $j^{\text{ème}}$ cellule de sortie

L'activation en sortie est notée o_j . Les cellules de sortie deviennent actives si leur degré d'activation (a_j) dépasse un seuil fixé (Θ_j). o_j

$$\text{aura pour valeur } \begin{cases} 0, & \text{pour } a_j \leq \Theta_j; \\ 1, & \text{pour } a_j > \Theta_j. \end{cases}$$

Règle d'apprentissage Le perceptron doit trouver l'ensemble des valeurs à donner aux synapses pour que les configurations d'entrée se traduisent par des réponses voulues à la sortie (cf. figure2). Pour cela, on utilise la règle d'apprentissage de Windrow-Hoff. Pour apprendre,

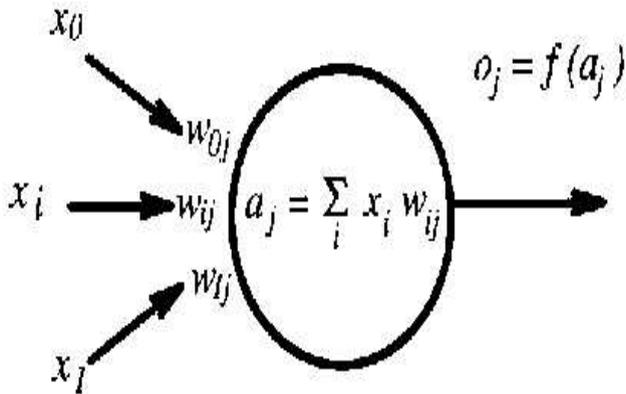


FIG. 2 -- Fonction d'apprentissage

le perceptron doit savoir qu'il a commis une erreur, et doit connaître la réponse qu'il aurait dû donner. De ce fait, on parle d'apprentissage supervisé. La règle d'apprentissage est locale dans le sens que chaque cellule de sortie apprend sans avoir besoin de connaître la réponse des autres cellules. La cellule ne modifie l'intensité de ses synapses (apprend) que lorsqu'elle se trompe. On va donner un ensemble de stimuli (intensité des synapses ou des connexions) au perceptron de façon arbitraire et on observe le résultat. Si le perceptron commet des erreurs, on lui permet de modifier l'intensité des connexions : les cellules du perceptron apprennent. Cette procédure est répétée jusqu'à ce que le perceptron soit capable de donner toutes les réponses correctes. Règle d'apprentissage de Windrow-Hoff: $w_{ij}^{(t+1)} = w_{ij}^{(t)} + \eta(t_j - o_j)x_i = w_{ij}^{(t)} + \Delta w_{ij}$ avec:

- x_i : valeur (0 ou 1) de la $i^{\text{ème}}$ cellule de la rétine
- o_j : réponse de la $j^{\text{ème}}$ cellule de sortie
- t_j : réponse théorique de la $j^{\text{ème}}$ cellule de sortie (0 ou 1)
- $w_{i,j}$: intensité de la connexion entre la $i^{\text{ème}}$ cellule de la rétine et la $j^{\text{ème}}$ cellule de sortie au temps t (il faut savoir que les valeurs de $w(0)$ sont choisies arbitrairement)
- η : constante positive dans le temps, comprise entre 0 et 1. Le problème du choix de η est souvent délicat. Sa valeur influence le temps d'apprentissage. Dans certains cas, η varie en fonction du temps : on commence avec une valeur élevée et puis diminue à chaque itération.

3.3 Séparabilité linéaire

On peut remarquer que le perceptron ne peut pas fonctionner correctement dans le cas représenté sur la figure 3, où les points représentatifs des deux classes sont regroupés dans des nuages qui s'interpénètrent, même légèrement. Dans ce cas, aucune droite séparatrice (aucun hyperplan) ne permettra de bien classer tous les exemples, c'est-à-dire de satisfaire la règle

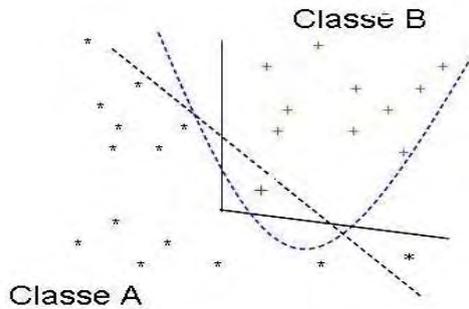


FIG. 3 -- Cas de non séparabilité linéaire

d'apprentissage, ce qui empêchera les coefficients de jamais se stabiliser. Le problème ainsi représenté est dit non linéairement séparable, ce qui peut être traité en utilisant l'approche de sélection de modèle linéaire par validation croisée (Shao, J., (1993). On peut montrer que la grande majorité des problèmes de classification dans des espaces d'entrée de forte dimensionnalité (qui sont les problèmes intéressants du point de vue pratique), et pour lesquels on possède un significatif d'exemples (donc pour lesquels on peut avoir l'espoir légitime de mener à bien l'apprentissage), sont non linéairement séparables, ce qui ruine apparemment la méthode que nous venons d'exposer.

On verra que ce problème peut être résolu par des astuces que nous citerons ultérieurement, notons au passage que la forte dimensionnalité n'est pas une condition nécessaire de la non séparabilité linéaire : la fonction booléenne OU exclusif, à deux entrées, n'est pas linéairement séparable, que nous citerons à titre d'exemple : Le cas le plus connu, ne possédant pas la propriété d'être linéairement séparable, est la fonction "XOR" (appelé aussi "ou exclusif"). Elle revient à associer une valeur de sortie pour deux cellules d'entrée de la manière suivante :

0 0 → 0
 1 0 → 1
 0 1 → 1
 1 1 → 0

Imaginons un perceptron avec deux cellules d'entrée et une cellule de sortie qui doit apprendre cette fonction. Il possède donc deux poids (intensité de connexion) w_1 et w_2 qui relient les cellules d'entrée à la sortie.

- Le fait d'associer les valeurs (1,0) d'entrées à '1' en sortie implique $w_1 > 0$.
- De même, le fait d'associer les valeurs (0,1) d'entrées à '1' en sortie implique $w_2 > 0$.
- Et donc, si nous mettons (1,1) d'entrées, le perceptron donnera toujours la réponse '1' ce qui est précisément le contraire de ce que l'on veut lui apprendre. Il est impossible de trouver les valeurs w_i correctes.

Pour le perceptron, deux catégories sont linéairement séparables si et seulement si les deux états de sortie peuvent être séparés par une droite (cf. figure ET).

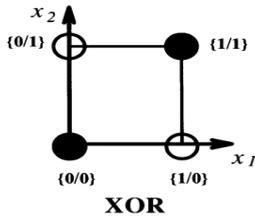


FIG. 4 -- Fonction XOR

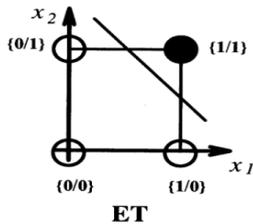


FIG. 5 -- Fonction ET

Par contre, la fonction logique ET est linéairement séparable car plus on "augmente" les entrées c'est-à-dire plus il y a des entrées à '1' et plus il est probable que la sortie passe à '1' ce qui n'était pas le cas de la fonction XOR. En effet, si une des entrées est à '1', la sortie est à '1' mais si les deux entrées sont à '1', la sortie va redescendre à '0'. Ces deux cas simples permettent de bien saisir le fait qu'une fonction soit linéairement séparable ou pas.

Pour rappel, le ET logique est :

0 0 → 0
 0 1 → 0
 1 0 → 0
 1 1 → 1

Le perceptron ne peut apprendre la fonction XOR justement parce qu'elle n'est pas linéairement séparable. En fait, il peut l'apprendre si elle est proprement codée : deux solutions existent.

1. Il suffit simplement d'utiliser trois cellules d'entrées. Le problème revient, ainsi, à apprendre une relation ternaire suivante :

0 0 0 → 0
 1 0 0 → 1

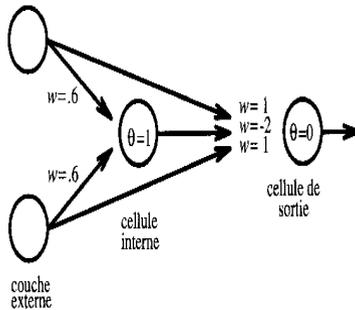


FIG. 6 -- Perceptron à plusieurs couches

0 1 0 → 1
 1 1 1 → 0

La troisième colonne d'entrée s'obtient par la multiplication des deux premières colonnes. Grâce à ce système, le perceptron peut résoudre le problème posé. Les poids qu'il faut associer sont $w_1 = 1, w_2 = 1$ et $w_3 = -2$. La difficulté en pratique est celui du codage du problème à résoudre plus que le problème de l'architecture du réseau.

- On crée un réseau avec plusieurs couches. Toujours concernant la fonction XOR, le perceptron sera composé de deux cellules d'entrées, une cellule cachée, une cellule de sortie et de connexions comme indiqué sur la figure 6.

On remarque que la cellule interne possède un seuil différent de zéro. Vérifions en calculant les quatre possibilités.

ACTIVATION DES CELLULES			
Entrées	Activation de la cellule interne ($\theta = 1$)	Etat de la cellule interne	Activation de la cellule de sortie ($\theta = 0$)
0 0	$(0 * 0.6) + (0 * 0.6) = 0$	0	$(0 * 1) + (0 * 1) + (0 * (-2)) = 0$
0 1	$(0 * 0.6) + (1 * 0.6) = 0.6$	0	$(0 * 1) + (1 * 1) + (0 * (-2)) = 1$
1 0	$(1 * 0.6) + (0 * 0.6) = 0.6$	0	$(1 * 1) + (0 * 1) + (0 * (-2)) = 1$
1 1	$(1 * 0.6) + (1 * 0.6) = 1.2$	1	$(1 * 1) + (1 * 1) + (1 * (-2)) = 0$

TAB. 2 -- Activation des cellules du perceptron.

La procédure la plus populaire d'apprentissage de la couche interne est connue sous le nom de rétropropagation ou de règle de Windrow-Hoff généralisée. Il en existe bien d'autres.

3.4 Perceptron multicouche

L'incapacité démontrée du perceptron à résoudre des problèmes intéressants entraîna, en son temps un ralentissement notable de la recherche sur le connexionnisme. Mais certains

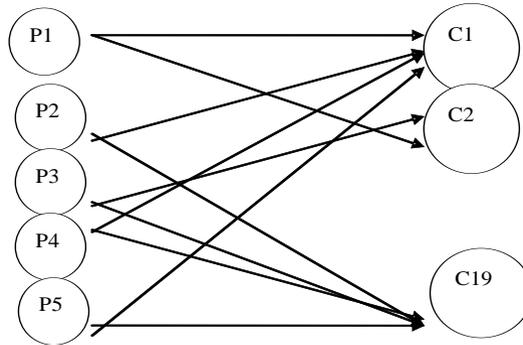


FIG. 7 – – Fonction d'activation ou d'affectation $O_k = f(Y_k)$

travaux démontrèrent dans le courant des années 70, que l'on pouvait quand même dépasser ces limitations en intercalant, entre la couche de cellule d'entrée et la couche de cellule de sortie, une ou plusieurs couches intermédiaires (appelées aussi couches cachées), l'information circulant d'une couche à la suivante. Dans ce cas on est certain de pouvoir approximer la sortie n'importe quelle fonction de l'espace d'entrée; mais si la théorie permet d'énoncer ce théorème d'existence, elle ne donne pas d'indication en revanche sur le nombre et la taille de ces couches cachées. Ce nombre devant être d'autant plus grand que cette fonction est plus irrégulière : puisque un neurone matérialise un hyperplan séparateur, une couche permet d'approximer une hypersurface par un ensemble de segments d'hyperplans, jouant le rôle de facette, des couches supplémentaires augmentant le degré de la fonction d'approximation. La question de l'apprentissage est restée un problème longtemps non résolu. En effet, si l'on peut appliquer à la dernière de poids les procédures applicables au perceptron monocouche, on est dans l'embarras pour les couches cachées car on ne connaît pas les sorties désirées des neurones, donc on est incapable de mettre en évidence l'erreur qui autoriserait la mise à jour des poids. La réponse a été apportée par plusieurs auteurs séparément, elle porte le nom générique d'apprentissage par rétropropagation de l'erreur (voir Bishop Christophe, M., (1995).

4 Application à l'assurance automobile

Etant donné un véhicule de la classe tarifaire dont le code est $xxxxx$. Nous avons établi (voir Tab.1) une classification qui a donné naissance à 19 classes d'agrégation, chacune compte une ou plusieurs classes du tarif actuel. Soit E l'ensemble de tous les codes tarifaires qui figurent dans l'échantillon. Nous avons $:xxxxx \in E, \exists$ un et un seul k tel que $xxxxx \in C_k$. Le problème qui se pose est de définir un réseau de neurones qui classe les (*) vecteurs $xxxxx$ représentant les classes de l'échantillon dans les 19 classes $C_k (k = 1...19)$. Pour une raison d'adaptation au problème traité, nous utilisons le perceptron à entrées réelles et sorties binaires. La figure 7 représente la situation.

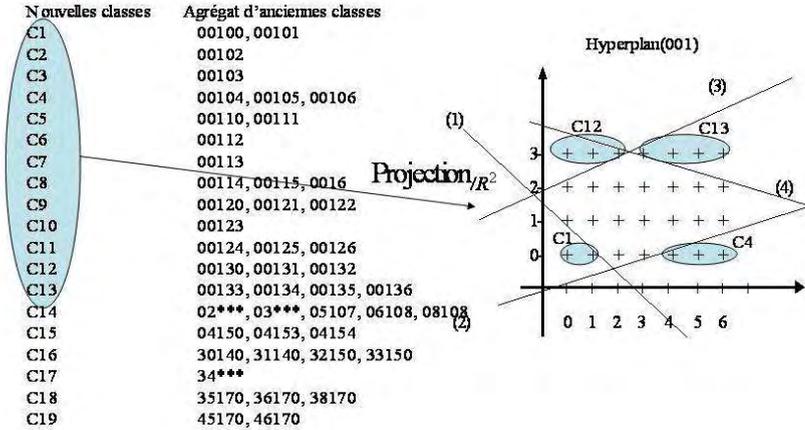


FIG. 8 -- Représentation géométrique des classes

La couche d'entrée est composée de 5 cellules ($P_i, i = 1 \dots 5$), chacune représente une modalité du code tarifaire. Chaque classe est représentée par un point de IR^5 .

Exemple : Pour la classe 05107, nous avons : $P_1 = P_4 = 0, P_2 = 5, P_3 = 1$ et $P_5 = 7$. La couche de sortie est composée de 19 cellules ($Y_k, k = 1 \dots 19$), une pour chaque classe C_k .

La sortie Y_k est égale à : $Y_k = \sum_{i=1}^5 w_{ik} P_i \quad k = 1 \dots 19$ On définit la fonction d'activation f : $O_k = f(Y_k) = \begin{cases} 0, & \text{si } Y_k \leq 0; \\ 1, & \text{sinon.} \end{cases}$ On a la propriété suivante : $xxxxx \in C_t \iff O_t = 1$ et $O_k = 0$ pour $k \neq t$ Le problème est de déterminer la matrice W .

4.1 Détermination des éléments de la matrice W

En premier lieu, nous allons nous intéresser aux 13 premières cellules de sorties. Celles-ci concernent les 28 classes des véhicules sans remorque de la zone Nord (les 001**). Nous allons les représenter géométriquement sur IR^2 (cf. figure8), puisque les trois premières composantes de leurs vecteurs sont identiques (ces points appartiennent à un sous espace vectoriel de dimension 2 de IR^5). Les autres vecteurs n'appartiennent pas à cet espace vectoriel : Chaque point sur le repère représente une classe, par exemple (2,3) représente la classe 00123. La droite (1) sépare l'ensemble C_1 des autres classes, la droite (2) fait la même chose Pour C_4 , la droite (3) pour C_{12} et la droite (4) pour C_{13} . Pour les neuf autres classes, ils n'existent pas de telles droites. Nous calculons par le biais de la règle d'apprentissage de Windrow-hoff, les éléments de la matrice W . L'utilisation d'une telle règle est motivée par la nature du problème traité.

- $C_1 = \{00100, 00101\}, w_{11} = w_{21} = -7, w_{31} = 2, w_{41} = -3$ et $w_{51} = -1$.

- $C_4 = \{00104,00105,00106\}, w_{14} = w_{24} = w_{44} = -7, w_{34} = -3$ et $w_{54} = 1$.
- $C_{12} = \{00130,00131,00132\}, w_{1,12} = w_{2,12} = -7, w_{3,12} = -6, w_{4,12} = 3$ et $w_{5,12} = -1$.
- $C_{13} = \{00133,00134,00135,00136\}, w_{1,13} = w_{2,13} = -7, w_{3,13} = -20, w_{4,13} = 6$ et $w_{5,13} = 1$.

Pour les autres classes, nous proposons la procédure adaptée que nous décrivons par un algorithme qui permet de résoudre le problème de la séparabilité linéaire :

4.2 Algorithme de calcul de la matrice W

DébutAlgorithme

$i = 6$; (initialement, il y a 5 cellules d'entées).

Pour $j = 1$ à 19 faire

Si C_j et E/C_j ne sont pas linéairement séparables alors :

Définir la cellule P_i fonction de P_1, P_2, \dots, P_5 telle que :

$$P_i \leq \alpha_i \forall x \in C_j$$

$P_i > \alpha_i \forall x \in E/C_j$ (cette cellule permet d'isoler les éléments de C_i des autres éléments de E).

Cette dernière rend C_i et E/C_i linéairement séparables, et nous calculons les poids.

FinSI

FinPour

FinAlgorithme.

L'application de l'algorithme donne les résultats suivants :

- $C_2 = \{00102\}$, on définit : $P_6 = P_4^2 + (P_5 - 2)^2$ Cette cellule mesure la distance entre le point $(0,2)$ et un autre de IR^2 . Cette dernière est nulle pour l'unique élément de C_2 , strictement positive pour tous les autres. Ainsi, C_2 et E/C_2 sont linéairement séparables, et nous calculons :
 $w_{12} = w_{22} = -7, w_{32} = 1, w_{42} = w_{52} = 0, w_{62} = -3$ et $w_{6k} = 0 \forall k \neq 2$ (i.e, la cellule P_6 n'intervient que pour C_2).
- $C_3 = \{00103\}$, on définit : $P_7 = P_4^2 + (P_5 - 3)^2$, et comme pour C_2 , nous calculons :
 $w_{13} = w_{23} = -7, w_{33} = 1, w_{43} = w_{53} = 0, w_{73} = -3$ et $w_{7k} = 0 \forall k \neq 3$.
- $C_5 = \{00110,00111\}$, on définit $P_8 = (P_4 - 1)^2 + (P_5 - 0,5)^2$:
 $w_{15} = w_{25} = -7, w_{35} = 1, w_{45} = w_{55} = 0, w_{85} = -3$ et $w_{8k} = 0 \forall k \neq 5$.
- $C_6 = 00112$, on définit : $P_9 = (P_4 - 1)^2 + (P_5 - 2)^2$:
 $w_{16} = w_{26} = -7, w_{36} = 1, w_{46} = w_{56} = 0, w_{96} = -3$ et $w_{9k} = 0 \forall k \neq 6$.
- $C_7 = 00113$, on définit : $P_{10} = (P_4 - 1)^2 + (P_5 - 3)^2$:
 $w_{17} = w_{27} = -7, w_{37} = 1, w_{47} = w_{57} = 0, w_{10,7} = -3$ et $w_{9k} = 0 \forall k \neq 7$.
- $C_8 = \{00114,00115,00116\}$, on définit : $P_{11} = (P_4 - 1)^2 + (\frac{P_5 - 5}{2})^2$:
 $w_{18} = w_{28} = -7, w_{38} = 1, w_{48} = w_{58} = 0, w_{11,8} = -3$ et $w_{11,k} = 0 \forall k \neq 8$.
- $C_9 = 00120,00121,00122$, on définit : $P_{12} = (p_4 - 2)^2 + (\frac{P_5 - 1}{2})^2$:
 $w_{19} = w_{29} = -7, w_{39} = 1, w_{49} = w_{59} = 0, w_{12,9} = -3$ et $w_{12,k} = 0 \forall k \neq 9$.
- $C_{10} = 00123$, on définit : $P_{13} = (p_4 - 2)^2 + (p_5 - 3)^2$:
 $w_{1,10} = w_{2,10} = -7, w_{3,10} = 1, w_{4,10} = w_{5,10} = 0, w_{13,10} = -3$ et $w_{13,k} = 0 \forall k \neq 10$.

- $C_{11} = \{00124,00125,00126\}$, on définit : $P_{14} = (p_4 - 2)^2 + (\frac{P_3 - 5}{2})^2$:
 $w_{1,11} = w_{2,11} = -7, w_{3,11} = 1, w_{4,11} = w_{5,11} = 0, w_{14,11} = -3$ et $w_{14,k} = 0 \forall k \neq 11$. A présent, intéressons-nous aux classes restantes. Nous utiliserons le même raisonnement que précédemment (nous définirons des cellules supplémentaires au besoin) :
- $C_{14} = \{02^{***}, 03^{***}, 05107, 06108, 08108\}$
 $w_{1,14} = -15, w_{2,14} = 7, w_{3,14} = 0, w_{4,14} = -6, w_{5,14} = -4$.
- $C_{15} = \{04150, 04153, 04154\}$
 $w_{1,15} = -1500, w_{2,15} = 100, w_{3,15} = -300, w_{4,15} = 0, w_{5,15} = -24$.
- $C_{16} = \{30140, 31140, 32150, 33150\}$
 $w_{1,16} = 4, w_{2,16} = -2, w_{3,16} = 0, w_{4,16} = -1, w_{5,16} = -7$.
- $C_{17} = \{34^{***}\}$. On définit : $P_{15} = (P_1 - 3)^2 + (P_2 - 4)^2$
 $w_{1,17} = 1, w_{2,17} = w_{3,17} = w_{4,17} = w_{5,17} = -4$ et $w_{15,k} = 0 \forall k \neq 15$.
- $C_{18} = \{35170, 36170, 38170\}$, on définit : $P_{16} = |36 - 10P_1 - P_2|$
 $w_{1,18} = 2, w_{2,18} = w_{4,18} = w_{5,18} = 0, w_{3,18} = 1, w_{16,18} = -2$ et $w_{16,k} = 0 \forall k \neq 16$.
- $C_{19} = \{45170, 46170\}$
 $w_{1,19} = 8, w_{2,19} = -3, w_{3,19} = w_{4,19} = w_{5,19} = 0$.

Pour notre problème, il s'agit d'affecter, en utilisant l'apprentissage par le réseau de neurones tel que nous avons conçu, un nouveau client, de caractéristiques connus à priori, à une classe de risque correspondante. Une prime de référence, lui est calculée en fonction de la nature du risque X de la classe d'affectation. Le calcul de cette prime pourra se faire, selon un principe de prime adéquat à la nature du risque considéré. Nous introduisons, dans le paragraphe suivant, différents principes de risque qui pourront être employés pour ce calcul. Cette démarche, que nous proposons, permet à la société d'assurance de mettre à jour, son système de tarification et à moindre coût, chaque fois que c'est nécessaire.

4.3 Principe de calcul de prime de base

Pour un risque X donné, variable aléatoire de loi de probabilité F_X , on associe une quantité $\prod(X)$, appelée prime de risque. Le calcul de cette prime est fondamental pour une compagnie d'assurance, car il représente un référentiel, pour le calcul de toute prime commerciale, notamment la prime de la Responsabilité Civile (RC), prime obligatoire et réglementée. Afin de donner les différents principes de calcul d'une prime de risque, nous commençons par donner les principales propriétés qui permettent le calcul d'une bonne prime de risque. Pour X, Y, Z trois risques arbitraires, nous avons les propriétés suivantes : (voir Rolski, T. , et al.(1998)):

Propriétés générales de prime de risque :

- $\forall \alpha \geq 0, \prod(\alpha) = \alpha$ (**non chargement**)
- $\forall \alpha \geq 0, \prod(\alpha X) = \alpha \prod(X)$ (**proportionnalité**)
- $\prod(X + Y) \leq \prod(X) + \prod(Y)$ (**subadditivité**) (**additivité, en cas d'égalité**)
- $\forall \alpha \geq 0, \prod(\alpha + X) = \alpha + \prod(X)$ (**Consistance**)
- si $X \leq_{st} Y$ alors $\prod(X) \leq \prod(Y)$ **préservation de l'ordre stochastique \leq_{st}**
- si pour tout $a \in [0,1]$, et pour tout Z ,

$$\begin{aligned} \prod(X) = \prod(Y) &\Rightarrow \prod(aF_X + (1 - a)F_Z) \\ &= \prod(aF_Y + (1 - a)F_Z) \quad (\text{compatibilité sous mixture}). \end{aligned}$$

Sur la base de ces propriétés, pour le calcul d'une bonne prime, il a été défini des principes de calcul de prime :

Principes de prime de risque

(A1) Principe de la valeur espérée : $\forall \alpha \geq 0, \Pi(X) = (1 + \alpha)E(X), (E(X) < \infty)$,

(A2) Principe de la variance : $\Pi(X) = E(X) + aVar(X)$,

(A3) Principe de l'écart type : $\Pi(X) = E(X) + a\sqrt{Var(X)}$,

(A4) Principe de la variance modifiée

$$\Pi(X) = \begin{cases} E(X) + a \frac{Var(X)}{E(X)}, & \text{si } E(X) > 0; \\ 0, & \text{si } E(X) = 0. \end{cases}$$

(A5) Principe de l'exponentiel:

$$\Pi(X) = \frac{\log E(e^{aX})}{a}$$

(A6) : Principe de la prime ajustée : Dans le cas d'un risque X qui se comporte suivant une loi de valeurs extrêmes, le principe suivant est proposé par (Wong , S. (1996)):

$$\Pi(X) = \int_u^\infty (1 - F_X(x))^{\frac{1}{p}} dx,$$

$p \geq 1$ est un paramètre de distorsion et u est un seuil convenablement déterminé.

(A7) : Principe de prime d'optimalité (cas du risque automobile): Dans (Boukhetala, K., (2001)). , nous proposons le principe suivant, qui cherche le calcul des primes optimales

$$\text{de périodes : (P 1) } \begin{cases} \text{Min}(Z) = \sum_{j \in H} C_j \pi_j \\ (1 - \alpha) \sum_{j \in H} nb_j \pi_j = \sum_{j \in H} S_j \\ \frac{\pi_j}{j} \leq \beta \left(\frac{\pi_{j-1}}{j-1} \right), j \in H - \{j_1\} \\ (1 - \alpha) \left(\frac{j}{j-1} \pi_{j-1} \right) \leq \pi_j \end{cases}$$

– $H = \{j_k\}_{k=1,2,\dots,h}$: Ensemble des périodes de contrats, par exemple: $j_h = 365$ jours. Dans le système considéré : $j_k = 3$ jours, 15 jours, 1 mois, 3 mois, 6 mois et une année.

– π : La prime à payer, pour la période j .

– α et β : Paramètres choisis par l'assureur, qui désignent respectivement le taux de chargement et le paramètre de solidarité ($0 < \alpha, \beta < 1$)

– nb_j : Nombre de sinistres observés sur la période j

– S_j : Montant de sinistre sur la période j

– C_j : Poids de préférence pour la période j .

4.4 Calcul de la prime: cas du principe de la moyenne

Pour le cas du risque automobile, le principe de la prime espérée est souvent utilisé, car empiriquement et grâce à la loi des grands nombres, il conserve le principe de solidarité entre les assurés d'une même classe de risque (avec une certaine neutralité au risque). Après avoir défini les classes d'agrégation, il nous reste à décrire la méthode utilisée pour le calcul de la prime. Soient alors :

C_m : le coût moyen d'un sinistre matériel (relatif à une classe ou à toute la population).

DUREE (JOUR)	NOMBRE DE CONTRATS	FREQUENCE
3	310	0,00284743
10	1420	0,01304308
20	369	0,00338936
31	13543	0,12439607
91	37094	0,34071829
183	36667	0,33679618
365	19467	0,17880959
Total	108870	1

TAB. 3 – – Réparation des contrats par durée.

DUREE	COEFFICIENT
3	0,01852222
10	0,06174074
20	0,12348148
31	0,1913963
91	0,56184075
183	1,12985558
365	2,25353708

TAB. 4 – – Coefficients de calcul de prime, correspondants à chaque durée.

F_m : la fréquence de sinistralité matérielle.

C_c : le coût moyen d'un sinistre corporel.

F_c : la fréquence de sinistralité corporelle.

D_m : la durée moyenne d'un contrat.

Pour une raison d'équilibre économique dans le portefeuille de l'assureur, la prime P doit vérifier la contrainte suivante :

$$\sum \text{des polices (cotisations)} \geq \sum \text{des coûts de sinistres.}$$

En égalant les deux membres de la formule, on obtient : $P = F_m \times C_m + F_c \times C_c$. Cette prime est applicable pour un contrat de durée moyenne D_m . Ce dernier est obtenu à partir de l'échantillon des contrats. La taille de notre échantillon de données est de l'ordre de 108870 contrats d'assurance. Le tableau suivant contient le nombre de contrats inscrits par durée :

Un simple calcul donne la prime pour toute durée donnée, en multipliant la prime P par le coefficient correspondant, comme c'est donné par le tableau : Suivant la nature des risques couverts, l'assureur peut également choisir, parmi les autres principes, celui qui est le plus approprié au calcul de sa prime.

4.5 Intervalle de confiance pour la prime : cas du principe de la moyenne

Pour une raison de simplicité pratique, nous préférons utiliser le principe de la moyenne. En pratique le calcul de la prime P est basé sur un échantillon de données sur la sinistralité. Le recours aux intervalles de confiances est souvent nécessaire afin de définir une région de confiance. Ainsi, et afin d'offrir aux décideurs une flexibilité dans le choix de la prime, pour aide à la tarification, nous définissons un intervalle I pour la prime pure centré autour de P , tel que:

$$I = [P_1, P_2], \text{ avec } P_1 = 0.95P \text{ et } P_2 = 1.05P$$

où le décideur aura la liberté de choisir une valeur appartenant à cette intervalle.

Les autres principes peuvent être utilisés, dans le cas où la loi de probabilité de la classe de risque considérée est identifiée à partir des données de la sinistralité engendrées par cette même classe. Pour le cas du principe de la prime ajustée (aversion au risque), un estimateur $\hat{\pi}_{u_n}$ de la prime est proposé dans (Necir, A., et Boukhetala, K.,2004), où le seuil u_n est convenablement déterminé. Un intervalle de confiance de cette prime est donné par le théorème suivant : (Necir, A., et Boukhetala, K., 2004) :

Pour $\gamma - 1/2 < 1/p \leq 1$, soit $k = k_n$ tel que pour $k \rightarrow \infty, k/n \rightarrow 0$,

$$\frac{(\frac{k}{n})^{-\frac{1}{p}} k^{\frac{1}{2}}}{Q(1-\frac{k}{n})} (\hat{\pi}_{u_n} - \pi_{u_n}) \rightarrow N(0, \sigma^2(p, \gamma)), \text{ quand } n \rightarrow \infty$$

où

$$\sigma^2(p, \gamma) \begin{cases} p^{-2} \gamma^{-2} \frac{\lambda^2 - \lambda + 1}{\lambda(\lambda + 1)}, & \text{pour } 1/p \neq \gamma; \\ 1, & \text{pour } 1/p = \gamma. \end{cases} \text{ avec } \lambda = 1/p - \gamma$$

γ est l'indice de la distribution à queue lourde associée au risque de valeurs extrêmes. Plusieurs estimateurs de ce paramètre sont proposés (Dekkers, A.L.M., Einmahl, J.H.J., and de Haan, L., (1989)), (Drees, H., (1995)), (Pickands, J., (1975.)). Le plus célèbre est l'estimateur de Hill (Hill, B., (1975)). Associé à un système d'information statistique de l'assurance automobile, tel que s'est décrit par (Aissani, D., et Boukhetala, K., (2006)), cette approche que nous proposons est mise en oeuvre sous forme d'un logiciel interactif RNPrime, présentée dans (Boukhetala, K., Laadjel, T. et Yahiaoui, N., (2001)). Ce logiciel est constitué de deux principaux modules, l'un est une base de données, traduisant le système d'information statistique considéré et l'autre est un module intégrant la méthode du réseau de neurones et la procédure de calcul de la prime pour chaque classe de risque. L'application contient également un module de simulation, pour aide à la tarification, en cas de manque de données réelles.

5 Conclusion

L'approche par réseaux de neurones, pour la mise à jour d'un système de tarification en assurance automobile, est proposée. L'estimation de la prime de base, par intervalle de confiance, présente un outil d'aide à la tarification. Un système d'information statistique, flexible et facilement accessible, doté d'une base de données fiable, est indispensable pour la validation statistique et économique du système de tarification proposé.

Références

- Aissani, D., et Boukhetala, K., (2006). *Stratégie de mise en place et de structuration du système d'information statistique algérien*. Revue CampusS, Vol.1, janvier 2006, pp 21-27.
- Boukhetala, K., Yahiaoui, M., et Laadjel, T., (2001). *Nouvelle segmentation en RC automobile par réseaux de neurones*. Rapport de recherche interne, USTHB.
- Boukhetala, K., (2001). *A linear programming Model for an optimal basic premium in car insurance* in Bulletin of the International Statistical Institute, Vol. III.
- Bishop Christopher, M., (1995) *Neural networks for pattern recognition*. Birmingham, Clarendon press Oxford.
- Dekkers, A.L.M., Einmahl, J.H.J., and de Haan, L., (1989). *A moment estimator for the index of an extreme-value distribution*. Annals of Statistics 17: 1833-1855.
- Drees, H., (1995). *Refined Pickands estimators of the extreme value index*. Annals of Statistics 23: 2059-2080.
- Habermanet, S., Pitacco, E., (1999). *Actuarial Models for disability insurance*. Edition Chapman .Hall.CRC.
- Haykin, S., (1994). *Neuronal Network. A comprehensive foundation*. Macmillan, New York.
- Hill, B., (1975). *A Simple Approach to Inference About Tail of a Distribution*. Ann. Statist., Vol.3, pp.1163-1174.
- Necir, A., et Boukhetala, K., (2004). *Estimating the risk-adjusted premium for the largest claims reinsurance covers*. Compstat, Proc. in Computational Statistics, Physica- Verlag, Heidelberg, NewYork.
- Partrat, C., et Besson, J-L, (2005). *Assurance non vie Modélisation, Simulation*. Edition Economica.
- Pickands, J., (1975). *Statistical inference using extreme order statistics*. Annals of Statistics, 3: 119-131.
- Rolski, T., Schmidli, H., Schmidt, V., Teugels, J., (1998). *Stochastic Processes for Insurance and Finance*. John Wiley Sons edition.
- Rosenblatt, R., (1958). *Principles of neurodynamics*. Spartan Books. New-York.
- Wong, S., (1996). *Premium calculation by Transforming the Layer Premium Density*. ASTIN Bulletin, 26,71-92.

Shao, J. (1993). *Linear model selection by cross-validation*. J. of the American Statistical Association, 88, 486-494.

Summary

The approach by artificial Neurons Network, for the updating of one tariff system in automotive insurance, is proposed. The estimated risk premium, at confidence intervals, presents a tool of help to the rating. A statistic information system, flexible and easily accessible, with a data base, is necessary for statistical and economical validation of the proposed rating system.

