

ESTIMATION DE QUANTITÉS SUBJECTIVES FLOUES PAR DES TECHNIQUES CONNEXIONNISTES. APPLICATION À L'ÉVALUATION DU CONFORT AUTOMOBILE.

Sylvain Millemann

Université de Technologie de Compiègne
U.R.A. CNRS 817
BP649 60206 Compiègne cedex

Kurtosis Ingénierie S. A.
27, av. de la Constellation
BP8295 95802 Cergy-Pontoise cedex

Pierre Scholl

Expertise métier liaison au sol
Renault S. A.
DIV / DSPV / 60865
Centre Technique d'Aubevoye
Le Parc de Gaillon
27940 Aubevoye

Résumé : En utilisant le formalisme de la logique floue dans le cadre des perceptrons multicouches, nous montrons dans cet article comment estimer le degré d'appartenance à une classe à partir d'observations subjectives incertaines, imprécises, voire conflictuelles. La méthode proposée est appliquée avec succès au problème de l'évaluation automatique du confort vibratoire automobile.

1. Introduction

De nombreux secteurs industriels ont encore recours aujourd'hui à l'expertise humaine pour le diagnostic d'un état de fonctionnement ou la reconnaissance d'un type d'événement et la quantification de son incidence. La cause peut être la difficulté d'automatiser certaines tâches, qui provient souvent de la nature même de celles-ci où la subjectivité joue un rôle important. Dans cet article, nous nous focalisons sur le problème de la qualification du confort, et plus précisément du confort vibratoire tactile automobile. La subjectivité de cette tâche se traduit par la variabilité du jugement :

- d'un individu à l'autre,
- en fonction du contexte extérieur (météorologie, milieu rural ou urbain, tendance du marché, ...),
- au cours du temps : changement d'état d'esprit.

Malgré ces variabilités dues à la nature subjective intrinsèque du confort, les experts sont capables de s'accorder pour situer de façon unanime un véhicule, au moins dans ses grandes tendances, par rapport à d'autres véhicules. Ceci nous autorise à penser qu'il est possible d'automatiser la construction d'un jugement global moyen sur le confort d'un véhicule, notamment en exploitant les avis conjoints de plusieurs experts.

Selon les experts Renault, la sensation globale de confort sur route est la résultante de sensations perçues localement au cours du temps. Or, la qualification de chacune de ces sensations est elle-même subjective : un obstacle localisé de type ralentisseur est une source d'inconfort plus ou moins importante selon les occupants d'un véhicule et d'un véhicule à l'autre. Ainsi, savoir représenter les informations incertaines et imprécises s'avère indispensable pour aborder le problème de la quantification automatique du confort automobile. Le formalisme de la logique floue, introduite par Zadeh [Zadeh 65], apparaît particulièrement bien adapté pour représenter ce type d'informations, habituellement gérées par l'être humain [Bouchon-Meunier 93]. Nous y consacrons la deuxième section de l'article, en se focalisant sur les outils nécessaires dans le cadre de cette étude.

Dans la troisième section, nous montrons comment construire un estimateur de degré d'appartenance à une classe floue grâce aux propriétés des perceptrons multicouches. On verra ainsi qu'un perceptron multicouche permet de fusionner des avis subjectifs, incertains, imprécis, voire conflictuels.

Enfin, la dernière section est consacrée à la mise en œuvre de la méthode proposée sur des données de confort automobile fournies par Renault.

2. Informations subjectives et floues

2.1. Types d'informations floues

Les informations floues, c'est-à-dire incertaines, imprécises, vagues, peu fiables, voire conflictuelles, se répartissent selon 3 catégories définies par Pal et Bezdek [Pal 94] : l'*incertitude probabiliste*, l'*incertitude résolutionnelle* et l'*incertitude floue*. Nous les présentons ci-dessous en s'attachant à faire référence au problème du confort automobile :

- L'*incertitude probabiliste* est liée au caractère aléatoire de l'occurrence d'un événement. Il s'agit par exemple de l'incertitude liée au profil *a priori* aléatoire de la route qu'un véhicule parcourt. Son importance est considérable puisque les aspérités de la route constituent la source même des principales sensations d'inconfort.
- L'*incertitude résolutionnelle* est le terme utilisé pour désigner l'incertitude liée aux imprécisions de mesures, c'est-à-dire aux limitations (souvent techniques) qui empêchent d'accéder aux informations de manière précise. On retrouve ce type d'incertitude lorsqu'un expert en confort hésite entre 2 phénomènes, qui lui procurent des sensations voisines.
- L'*incertitude floue* désigne la part d'imprécision due au codage utilisé pour représenter l'information. Les termes utilisés dans ce codage définissent des ensembles dont les frontières ne sont pas strictes, par définition. Les termes employés par les experts pour qualifier le confort automobile entrent dans ce cadre. Par exemple, le terme générique de *pompage*, qui désigne un mouvement vertical basse fréquence de la caisse, ne définit pas un ensemble aux frontières strictes : un faible *tassement* peut en effet souvent être assimilé à un fort *pompage*.

2.2. Formalisme de la logique floue

Nous présentons ci-dessous quelques aspects de la théorie de la logique floue, fondamentaux pour la construction du système de détection/estimation à la section 3. Le lecteur trouvera une présentation complète dans [Dubois 80].

Principes

Désignons par X l'univers de référence et plaçons-nous d'abord dans la théorie des ensembles classiques, c'est-à-dire nets. Alors, tout sous-ensemble A de X est totalement défini par sa fonction caractéristique μ_A de X vers l'ensemble d'évaluation $\{0, 1\}$:

$$\mu_A(x) = \begin{cases} 1 & \text{ssi } x \in A \\ 0 & \text{ssi } x \notin A \end{cases} \quad (1)$$

Lorsque l'ensemble d'évaluation est le continuum $[0, 1]$, A est appelé **sous-ensemble flou** de X (et noté \tilde{A}) et $\mu_{\tilde{A}}$ est sa **fonction d'appartenance**. On a alors :

$$\tilde{A} = \left\{ (x, \mu_{\tilde{A}}(x)), x \in X \right\} \quad (2)$$

Dans le souci d'alléger les notations, l'ensemble flou \tilde{A} sera noté A quand aucune ambiguïté n'est possible.

Au cours d'un essai sur route, l'expert en confort s'exprime à l'aide de variables linguistiques, en qualifiant les défauts ressentis de *très* critiquables, ou *faiblement* prononcés, etc. Chaque phénomène est ainsi naturellement gradué en fonction de son importance pour l'occupant. Un type d'inconfort peut alors se définir comme un sous-ensemble flou de l'ensemble net des sensations potentielles de l'occupant.

En désignant par z_i une zone temporelle d'un essai sur route et par k une classe d'inconfort (par exemple la *pompage*), on ne cherche pas à dire que z_i appartient ou n'appartient pas à la classe k , mais plutôt à savoir dans quelle mesure z_i appartient à k . On cherche ainsi par exemple à quantifier la présence du *pompage* dans la zone temporelle z_i .

Classification

Dans le cadre de la théorie des ensembles nets, établir une classification de l'univers de référence X vers un ensemble de classes nettes $\Omega = \{\omega_1, \dots, \omega_m\}$ revient à construire une règle de décision d permettant d'affecter tout élément x de X à un seul élément ω_i de Ω :

$$d : \begin{cases} X \rightarrow \Omega \\ x \mapsto d(x) \end{cases} \quad (3)$$

Chaque ω_i peut alors être vu *a posteriori* comme une partie de X , et $d^{-1}(\Omega)$ comme une partition de X . On a alors, pour tout x de X :

$$\sum_{i=1}^m \mu_{\omega_i}(x) = \mu_{d(x)}(x) = 1 \quad (4)$$

On se focalise ici sur la classification supervisée où la règle de décision d est construite à partir d'une base de couples (*donnée, classe d'appartenance*) déjà connus. Cette base est séparée en 2 ensembles : un ensemble sur lequel l'apprentissage est effectué (base d'apprentissage) et un ensemble permettant d'estimer la performance de la règle d obtenue, en dénombrant les éléments bien classés (base de test). En effet, dans le cadre usuel des ensembles nets, il est rare de remettre en cause la fiabilité de la base de données (ceci est en outre rarement concevable car il s'agit souvent de la seule connaissance disponible sur le système hormis pour les systèmes évolutifs qui ne font pas partie du cadre de cette étude).

L'extension de la classification au cadre des ensembles flous consiste à construire une règle de décision d_f qui n'effectue plus un choix tranché entre l'ensemble des classes, mais qui estime le degré d'appartenance de x à chacune des classes. On peut le formaliser comme suit :

$$d_f : \begin{cases} X \rightarrow [0, 1]^m \\ x \mapsto d_f(x) = \{\mu_1(x), \dots, \mu_m(x)\} \end{cases} \quad (5)$$

Dans ce contexte, on distingue 2 cas :

- $d^{-1}(\Omega)$ ne constitue pas une partition de X :
Ceci peut s'interpréter en disant que les degrés d'appartenance de x à 2 classes ω_i et ω_j ne sont pas liés. Si les classes sont définies par des phénomènes confort, alors la relation :

$$\mu_{\omega_i}(z) = \mu_{\omega_j}(z) = 0.9 \quad (6)$$

signifie par exemple que les 2 phénomènes ω_i et ω_j sont fortement prononcés simultanément.

- $d^{-1}(\Omega)$ constitue une partition de X :
Dans le cas flou, on a alors pour tout z de X , selon la définition de Ruspini [Ruspini 69] :

$$\sum_{i=1}^m \mu_{\omega_i}(z) = 1 \quad (7)$$

Cette relation n'est pas vérifiée lorsque les classes ω_i désignent les différents phénomènes confort. Par conséquent, chaque phénomène sera étudié individuellement.

Contrairement au cas net, les couples (*donnée, classe d'appartenance*) utilisés pour un apprentissage supervisé ne peuvent plus être considérés comme totalement fiable, en raison de la nature même des phénomènes traités. La notion même de fiabilité perd en outre un peu de son sens, notamment dans le

cas où une classe est définie à partir d'une expertise humaine, c'est le cas par exemple où l'appartenance à cette classe est variable d'un expert à l'autre. Parler de bon ou mauvais classement n'a donc pas beaucoup de sens en terme de classification floue. L'objectif réside seulement dans l'obtention d'une bonne adéquation entre l'appartenance à une classe notée par des experts et la règle de décision estimée.

2.3. Outils adaptés à l'expertise humaine

Les différents types d'informations floues se traduisent, comme il a déjà été précisé, par des imprécisions, des incertitudes, voire des conflits. Les causes diverses et multiples qui interdisent de disposer d'une information nette (cf. § 1.), rendent difficile une quantification du degré de flou. Or, cette quantification s'avère capitale dans la perspective de savoir juger la qualité d'une classification floue. Pour chaque type de défaut confort, il est important de savoir évaluer la difficulté à le reconnaître et le quantifier.

Outre les mesures de flou, analogues à des distances ou encore à des entropies, nous proposons ci-après une mesure, appelée *cardinalité relative* permettant de comparer deux classes floues à travers les résultats d'une règle de décision.

Mesure de flou

De nombreuses mesures de flou sont proposées dans la littérature. Pal et Bezdek en proposent une synthèse dans [Pal 94]. Ces mesures sont rangées selon leurs types, dans différentes classes : classe de Ebanks, classe de Loo, ou encore de Pal et Bezdek pour les plus récentes. Nous nous focalisons sur l'une d'elles, très utilisée : la mesure de flou H_{Ka} de Kaufmann définie dans [Kaufmann 73]. Elle s'écrit, pour l'ensemble A :

$$H_{Ka}(A, q) = \frac{2}{n^{1/q}} \left[\sum_{x \in X} \left| \mu_A(x) - \mu_{A_{1/2}}(x) \right|^q \right]^{1/q} \quad (8)$$

où $q \in [1, +\infty[$, n désigne le nombre d'éléments de l'univers de référence X et $A_{1/2}$ l'ensemble net le plus proche de A , c'est-à-dire son α -coupe de niveau $1/2$ définie par :

$$A_{1/2} = \{x \in X \mid \mu_A(x) \geq 1/2\} \quad (9)$$

H_{Ka} est en fait une distance de Minkowski calculée entre l'ensemble flou A et son ensemble net le plus proche. Ainsi, plus les degrés d'appartenance de A sont distants (respectivement proches) de $1/2$, c'est-à-dire proches (respectivement distants) de 0 ou de 1, et plus H_{Ka} est faible (respectivement élevé).

Cardinalités relatives

Évaluer la performance d'une classification, dans le cas net revient à comptabiliser la proportion d'éléments bien classés. Dans le cadre plus général de la logique floue, cette démarche n'est plus possible et n'a guère de sens. Il est néanmoins possible de rester dans le même esprit en mesurant une quantité d'appartenance moyenne de l'ensemble des éléments à une classe. En normalisant par rapport à la quan-

tité d'appartenance globale, on dispose ainsi d'une mesure prenant ses valeurs entre 0 et 1. Nous appelons *cardinalité relative* de la classe j par rapport à la classe i la quantité $CR_i(\omega_j)$:

$$CR_i(\omega_j) = \frac{\sum_{\substack{x \in \omega_i \\ \text{expert}}} \mu_j(x)}{\sum_{\substack{x \in \omega_i \\ \text{expert}}} \sum_{k=1}^m \mu_k(x)} \quad (10)$$

où $\sum_{\substack{x \in \omega_i \\ \text{expert}}}$ signifie que la somme est constituée des observations étiquetées ω_i par l'expert. Il s'agit donc, pour l'ensemble des éléments étiquetés dans ω_i , de la quantité d'appartenance moyenne à la classe ω_j . On obtient ainsi l'équivalent des matrices de confusion, usuelles en classification nette. Les termes diagonaux $CR_i(\omega_i)$ doivent retenir l'attention. En effet :

- Lorsque $CR_i(\omega_i)$ est voisin de 1, on en déduit que la classe ω_i présente des frontières nettes mises en évidence par la classification.
- Par contre, une cardinalité $CR_i(\omega_i)$ distante de 1 traduit la nature floue intrinsèque de la classe ω_i . Les cardinalités relatives fournissent une information pertinente sur la nature des classes, en terme de quantité de flou.

La Figure 1 montre une illustration graphique possible des cardinalités relatives, dans le cas de 2 classes. Le trait plein est utilisé pour l'ensemble d'apprentissage, le trait pointillé pour l'ensemble test. Les cardinalités relatives sont d'autant plus proches des valeurs extrêmes 0 et 1 que les classes sont nettes et bien reconnues. Par ailleurs, la proximité des valeurs obtenues pour base d'apprentissage et base de test permet de valider, dans une certaine mesure, le découpage de la base de données.

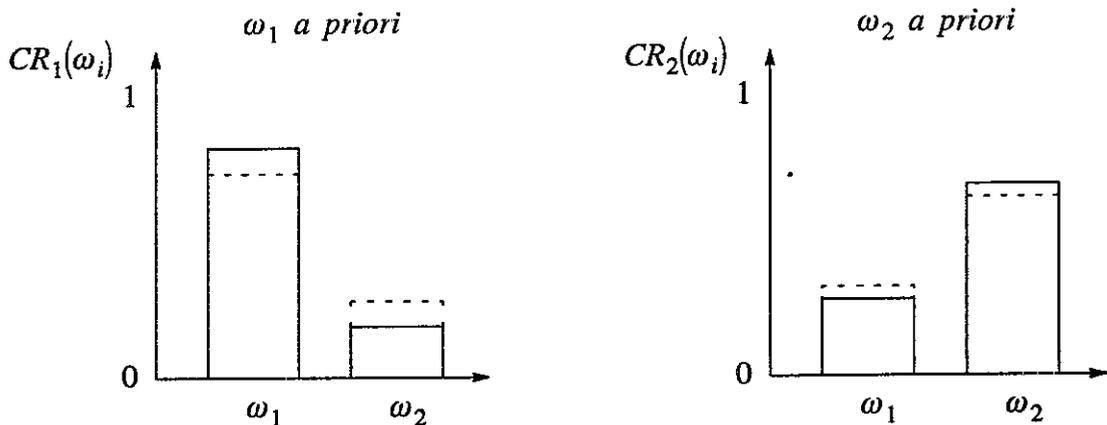


Figure 1
Cardinalités relatives estimées dans le cas de 2 classes

Dans l'application à l'évaluation du confort automobile, chaque classe d'inconfort est définie par les sensations de l'expert. Les cardinalités relatives nous donneront donc une information sur le degré de facilité avec lequel les experts reconnaissent un type d'inconfort et savent le distinguer des autres.

Dans le cadre de la classification supervisée, une bonne cohérence entre les cardinalités relatives estimées sur la base d'apprentissage et les cardinalités relatives estimées sur la base de test permet de valider l'apprentissage (et dans une certaine mesure le découpage de la base de données).

3. Estimation connexionniste de fonctions d'appartenance floue

Résoudre un problème de classification dans le cadre des ensembles flous revient à construire un système permettant d'estimer les degrés d'appartenance de tout élément de l'univers de référence, à chacune des classes ω_i . Pour effectuer cette tâche, nous avons choisi, compte tenu de leurs propriétés, d'utiliser les Perceptrons MultiCouche (PMC). La littérature est riche sur ce sujet (cf. par exemple [Rumelhart 86], [Hertz 91] ou encore [Bottou 91]). C'est pourquoi nous nous contenterons d'en faire une présentation succincte, en abordant uniquement les spécificités qui en font un outil particulièrement adapté à notre problème.

3.1. Le Perceptron MultiCouche

Dans sa forme la plus classique, le PMC fait partie de la classe des réseaux de neurones statiques, implémentant le plus souvent des transformations non linéaires du type :

$$\phi : \begin{cases} \mathbb{R}^n \rightarrow I \\ x \mapsto \phi(x) \end{cases} \quad (11)$$

Le PMC est notamment bien adapté pour les problèmes de classification supervisée pour lesquels il est capable d'apporter des solutions non linéaires. Nous le présentons dans ce contexte.

Un PMC est constitué de 3 types de couches (cf. Figure 2) :

- une couche d'entrées où sont présentés les vecteurs caractérisant les observations à classer,
- une couche de sortie dont le rôle est de délivrer les vecteurs codant la décision,
- une ou plusieurs couches cachées au sein desquelles se propage l'information.

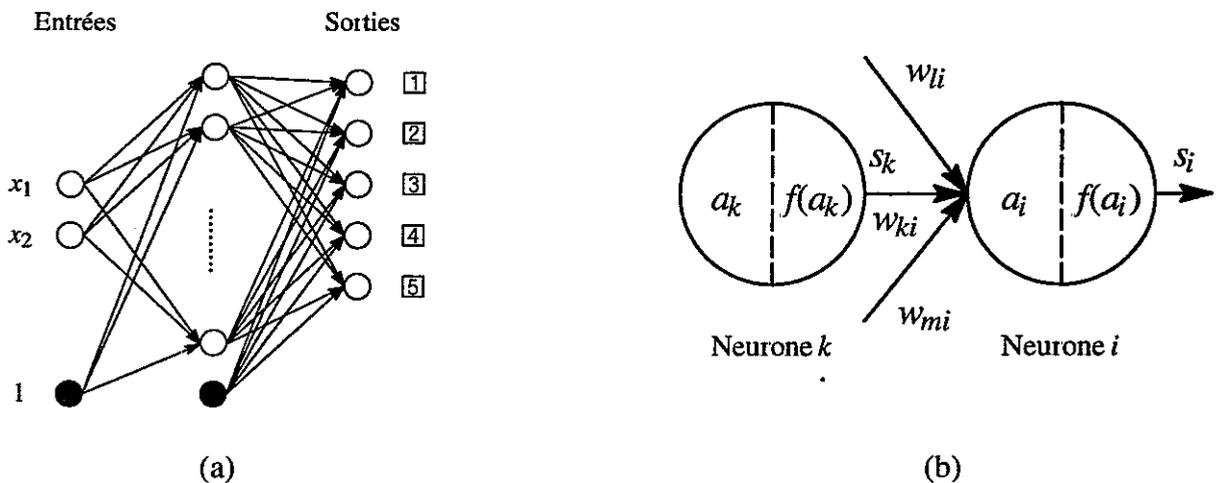


Figure 2
Perceptron multicouche

En désignant par w_{ki} le poids de la connexion reliant les neurones k et i (cf. Figure 2b), la propagation s'effectue, à l'échelle du neurone i , comme suit :

- Calcul de l'activation :

$$a_i = \sum_{k \in \text{Amont}(i)} w_{ki} s_k \quad (12)$$

- Calcul de la sortie :

$$s_i = f(a_i) \quad (13)$$

f désignant la fonction de transfert du neurone i , souvent définie par une sigmoïde.

La phase d'apprentissage du réseau, c'est-à-dire la construction de la règle de décision, consiste à ajuster progressivement les poids w_{ki} des connexions en optimisant un critère sur la couche de sortie. Le cas le plus répandu est l'utilisation de l'algorithme de *rétropropagation du gradient* (cf. [Rumelhart 86]) pour minimiser l'erreur quadratique entre les sorties désirées s^* et les sorties obtenues s :

$$J_{\phi_w} = \sum_{\substack{k \in \mathcal{Y} \\ x \in X}} (s_k^*(x) - s_k(x))^2 = \sum_{\substack{k \in \mathcal{Y} \\ x \in X}} (s_k^*(x) - \phi_{w,k}(x))^2 \quad (14)$$

\mathcal{Y} désignant l'ensemble des neurones de la couche de sortie.

Propriétés

Trois propriétés établies sur les PMC font de ceux-ci un outil très utilisé. Nous les abordons ci-dessous.

- Lien avec l'analyse de données :

Bourlard [Bourlard 88] montre qu'un PMC peut être utilisé pour effectuer une analyse en composantes principales. Par ailleurs, Gallinari dans le cas linéaire [Gallinari 91] et Webb dans le cas non linéaire sous certaines contraintes [Webb 90], ont montré l'équivalence entre un PMC contenant une seule couche cachée et l'analyse discriminante.

- Approximateur universel :

Les PMC sont des approximateurs universels. De nombreuses équipes ont travaillé sur ce sujet. On peut notamment citer Cybenko [Cybenko 89] et Hornik [Kornik 89] qui ont montré parallèlement qu'il est possible d'approcher toute fonction continue à l'aide d'un PMC comportant 2 couches de connexions (c'est-à-dire une seule couche de neurones cachée), à condition d'utiliser un nombre suffisant de neurones. Ce résultat intéressant doit cependant être modéré par le manque de connaissance sur l'incidence que peut avoir le choix de 2, voire de 3 couches de neurones cachées au lieu d'une seule. On sait par exemple que certains problèmes nécessitent un nombre exponentiel de neurones si le réseau ne comporte qu'une seule couche cachée, alors qu'ils en nécessitent un nombre polynomial lorsque le réseau est construit avec 2 couches cachées.

- **Estimateur bayésien :**

Cette propriété est d'une importance considérable compte tenu du grand nombre d'applications du domaine de la classification où les PMC sont utilisés. Elle le sera notamment dans notre application en confort automobile. C'est pourquoi nous la présentons plus en détail dans la section 3.2 suivante.

Mises en œuvre

Trois grands problèmes restent néanmoins ouverts aujourd'hui sur l'aspect pratique de l'utilisation des PMC : savoir diminuer la complexité de l'apprentissage, savoir dimensionner le PMC en fonction de l'application, et savoir contrôler son pouvoir de généralisation.

- **Complexité d'apprentissage :**

Des 3 points cités ci-dessus, l'optimisation des algorithmes d'apprentissage mis en œuvre sur les PMC est sans doute le plus difficile. D'une part, il a été prouvé que la détermination des poids des connexions optimaux, pour une base d'apprentissage donnée et pour une taille de PMC donnée, est un problème NP-complet [Judd 90]. Les propositions faites pour diminuer la complexité sont souvent très spécifiques à un type de données et demandent des réglages délicats.

- **Dimensions du PMC :**

Un PMC sous-dimensionné est incapable d'apprendre la base de données pour laquelle on l'utilise. C'est le cas lorsque la couche cachée contient trop peu de neurones. En revanche, surdimensionner un PMC peut conduire à un apprentissage par cœur de la base, c'est-à-dire à une solution qui ne serait pas applicable à de nouveaux éléments, non issus de la base d'apprentissage. Il est donc important de bien savoir doser la taille d'un PMC. Plusieurs méthodes sont proposées dans la littérature :

- Le PMC peut être dimensionné en s'inspirant des modèles statistiques [Ciampi 95], comme les arbres de décisions (cf. [Sethi 90], [Millemann 91] ou encore [Chabanon 92]).
- Une solution dite ascendante, consiste à construire progressivement le réseau, neurone par neurone, voire couche par couche, afin de pouvoir garder la meilleure architecture relativement à un critère donné (cf. par exemple l'algorithme de cascade de corrélation de Fahlman [Fahlman 90]). Cette méthode vérifie itérativement que l'ajout d'un neurone (ou d'une couche) ne conduit pas à un résultat moins bon sur une base de test. Lengellé et Dencœur [Lengellé 92] proposent, dans le cadre de la classification supervisée, de construire le réseau en améliorant le pouvoir de séparabilité linéaire d'une couche à l'autre. Dans ce cadre, le rôle des neurones cachés est de coder les entrées en formant des nuages de points de plus en plus compacts.
- Une solution dite descendante utilise la démarche inverse. Il s'agit alors de surdimensionner le réseau de neurones puis de l'élaguer progressivement jusqu'à l'obtention de la valeur optimale du critère. Cette méthode, développée dans [Le Cun 90], nécessite toutefois de savoir surdimensionner le réseau *a priori*.
- Une solution mixte la première et la troisième méthode [Mtimet 94].

- **Généralisation :**

On juge les performances d'une classification aux résultats obtenus sur une base de test représentative de l'ensemble des données. En effet, le classifieur est d'autant meilleur qu'il est capable de généraliser son comportement sur des données non apprises. Un bon pouvoir de généralisation n'est envisageable qu'avec un bon compromis entre la taille du réseau et la taille de la base d'apprentissage. Lorsque la taille de la base est fixée, optimiser le pouvoir de généralisation nous ramène au problème du dimensionnement du réseau, et au choix d'un critère d'arrêt judicieux pour l'apprentissage.

Que ce soit pour optimiser la taille du PMC ou pour contrôler de façon efficace l'arrêt de l'apprentissage, une pratique efficace et répandue est la validation croisée. Elle consiste tout d'abord à séparer la base de données en 2 ensembles : base d'apprentissage et base de test. À chaque itération (ajout d'un neurone ou mise à jour des poids des connexions), l'évolution du critère à optimiser (par exemple l'erreur quadratique en sortie du PMC), est surveillée sur la base de test. Dès que la tendance de celle-ci n'est plus à la baisse, l'apprentissage est arrêté.

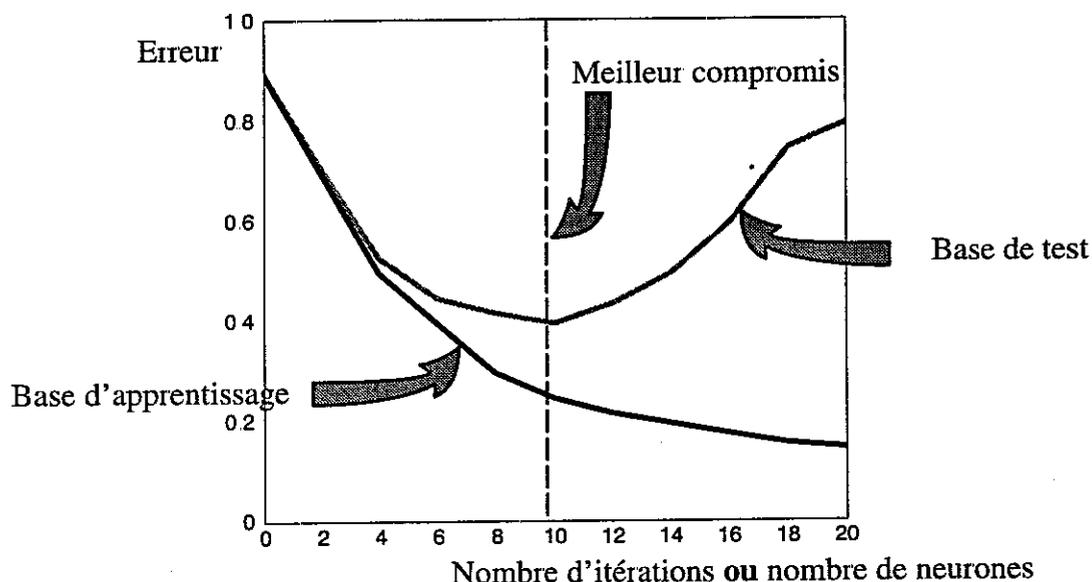


Figure 3

Validation croisée sur le nombre d'itérations ou sur le nombre de neurones

3.2. Estimateur bayésien

Nous développons ci-dessous un raisonnement conduisant à la propriété d'estimateur bayésien pour les PMC. D'autres approches sont proposées dans la littérature (cf. [Ruck 90] ou encore [Richard 91]). Plaçons-nous à nouveau dans le cadre de la classification nette, où l'on cherche à construire une règle de décision d telle que présentée par (3), permettant de classer tout élément x de $X = \mathbb{R}^n$, dans l'ensemble des m classes $\Omega = \{\omega_1, \dots, \omega_m\}$. Désignons par $P(\omega_i | x)$ la probabilité *a posteriori* de ω_i connaissant x . Lorsque les m classes forment une partition de \mathbb{R}^n , la règle de décision de Bayes est la règle optimale définie par :

$$d_B = \omega_j \tag{15}$$

avec

$$\forall i \neq j \quad P(\omega_i | x) \leq P(\omega_j | x) \tag{16}$$

On affecte ainsi x à la classe dont la probabilité conditionnelle est la plus forte. Soit à présent un PMC possédant n neurones sur la couche d'entrée et m neurones sur la couche de sortie. Alors, la classe d'appartenance de tout vecteur $x \in \mathbb{R}^n$ de la base d'apprentissage est codée par un vecteur binaire s^* de façon que :

$$x \in \omega_i \Rightarrow \begin{cases} s_i^* = 1 \\ s_j^* = 0 \end{cases} \quad j = 1, \dots, m \quad j \neq i \quad . \quad (17)$$

La généralisation du critère (14) à un nombre non forcément fini d'éléments, peut s'écrire, en désignant par E l'espérance mathématique :

$$J_{\phi_w} = E(|s^* - s|^2) \quad (18)$$

La minimisation de ce critère conduit à estimer, en sortie du réseau, l'espérance $E(s^* | x)$, ce qui s'écrit, compte tenu du codage (17) utilisé, au niveau du neurone j :

$$E(s_j^* | x) = 0 \cdot P(s_j^* = 0 | x) + 1 \cdot P(s_j^* = 1 | x) \quad (19)$$

c'est-à-dire :

$$E(s_j^* | x) = P(\omega_j | x) \quad (20)$$

La propriété d'approximateur universel permet donc d'en déduire, à condition que la base d'apprentissage soit suffisamment représentative des données à classer, que le PMC fournit une approximation de la règle de décision optimale de Bayes, en estimant les probabilités *a posteriori* $P(\omega_i | x)$:

$$\forall i \in \{1, \dots, m\} \quad s_i(x) = \hat{P}(\omega_i | x) \quad (21)$$

3.3. Estimation d'appartenance floue

Cette partie est consacrée à la transposition de la propriété d'estimateur bayésien dans le contexte de la logique floue (détaillée aussi dans [Millemann 95] et [Millemann 96]). Nous introduisons tout d'abord la notion de probabilité subjective.

Probabilité subjective

Lorsque ce sont des mesures (objectives) qui définissent des classes d'événements, la notion de probabilité est directement rattachée à la fréquence d'occurrence d'un événement. Cette probabilité peut être qualifiée d'objective car elle ne dépend d'aucun facteur humain, et ne fait pas intervenir la notion de doute.

Lorsque les classes d'événements sont définies à partir d'opinions humaines, l'information accessible n'est alors plus complètement fiable ni précise. Les discordances, les doutes et les imprécisions du jugement nous interdisent alors de parler de probabilité. Cependant, même si on ne peut en effet plus fournir une fréquence d'apparition d'un événement (ce qui n'a d'ailleurs pas toujours de sens), on peut néanmoins connaître cette fréquence à travers les jugements subjectifs des experts humains. On peut alors parler de la probabilité que le diagnostic subjectif des experts soit en faveur de tel événement. On utilise alors le terme de probabilité subjective, introduit par Shafer [Shafer 76] et repris notamment par

Zadeh [Zadeh 84]. On prendra garde de ne pas confondre ces probabilités avec les probabilités *a priori* pour lesquelles on utilise parfois aussi ce terme de probabilité subjective [Wonnacott 91].

Étudions à présent les liens qui unissent probabilité subjective et fonction d'appartenance floue. Pour fixer les idées, considérons l'exemple simple de l'ovale, issu de [Kosko 92]. Dans cet exemple, on pose la question "La forme de la Figure 4 appartient-elle à la classe des ovales ?".

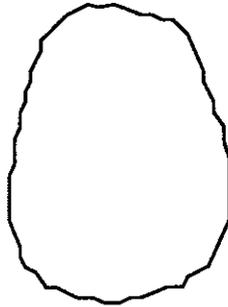


Figure 4
Un ovale ?

La réponse est forcément très variable en fonction de la personne questionnée. Ainsi, dire que la forme de la Figure 4 appartient à la classe des ovales avec la probabilité P_0 n'a pas de sens. En revanche, la probabilité (subjective) que les individus se prononcent pour la réponse oui est tout à fait accessible. Une valeur de 80% signifierait alors que cette forme est un ovale pour 80% des individus. On dispose ainsi d'un moyen de quantifier l'appartenance de la forme à la classe des ovales, c'est-à-dire d'un degré d'appartenance à cette classe. En effet, ce qui importe dans la question "La forme appartient-elle à la classe des ovales ?" est plus le consensus des personnes interrogées qu'une mesure objective de l'appartenance. Notons que ce type de raisonnement, applicable à tout diagnostic humain (et donc à l'évaluation du confort automobile), a déjà été utilisé, notamment dans le domaine des sciences sociales [Nowakowska 77].

Estimation des degrés d'appartenance à l'aide d'un PMC

Dans le cadre de la classification floue, c'est à partir des avis des experts humains que l'apprentissage du PMC doit s'effectuer. Il est donc nécessaire de savoir coder ces diagnostics subjectifs sur la couche de sortie du PMC. Nous proposons pour cela d'utiliser un codage binaire, voisin de (18), c'est-à-dire de garder un codage net, pour les commentaires flous des experts. L'appartenance de x à la classe ω_i , selon un expert est formulée par $x \underset{\text{expert}}{\in} \omega_i$. Le codage utilisé sur la dernière couche pour représenter les sorties désirées devient :

$$x \underset{\text{expert}}{\in} \omega_i \Rightarrow \begin{cases} s_i^* = 1 \\ s_j^* = 0 \quad j = 1, \dots, m \quad j \neq i \end{cases} \quad (22)$$

Comme il a été montré au paragraphe 3.2, les neurones situés sur la couche de sortie du PMC estiment les probabilités *a posteriori*. Comme cette fois, l'apprentissage est fondé sur des avis subjectifs, les $\hat{P}(\omega_i | x)$ estimées sont les probabilités subjectives *a posteriori* que les experts se prononcent en faveur

de la classe ω_i . Compte tenu en outre du lien établi entre probabilité subjective et fonction d'appartenance floue au début de cette section, la valeur fournie par le neurone i de la couche de sortie du PMC peut être interprétée comme une estimation du degré d'appartenance à la classe ω_i :

$$\forall i \in \{1, \dots, m\} \quad s_i(x) = \hat{P}(\omega_i | x) = \hat{\mu}_i(x) \quad (23)$$

Bien sûr, ce résultat n'est valable que sous la contrainte que les classes ω_i impliquées forment une partition floue, c'est-à-dire :

$$\sum_{i=1}^m P(\omega_i | x) = \sum_{i=1}^m \mu_i(x) = 1 \quad (24)$$

Minimisation d'une mesure de flou

Le critère quadratique (14), optimisé lors de l'apprentissage, s'écrit pour chaque neurone i de la couche de sortie :

$$J_{\phi,i}(x) = \sum_{x \in A} (s_i^*(x) - s_i(x))^2 = \sum_{x \in A} (s_i^*(x) - \phi_{w,i}(x))^2 \quad (25)$$

qui devient désormais :

$$J_{\phi,i}(x) = \sum_{x \in A} (s_i^*(x) - \hat{\mu}_i(x))^2 \quad (26)$$

Ce critère mesure une distance entre l'appartenance de x à la classe ω_i telle que les experts peuvent la fournir ($s_i^*(x)$) et celle estimée par le PMC ($\hat{\mu}_i(x)$). Cette mesure s'avère être très semblable à la mesure de flou de Kaufmann (8). Au lieu de mesurer la distance moyenne des degrés d'appartenance à l'ensemble net le plus proche, on mesure cette fois la distance par rapport à l'ensemble net défini par le codage des commentaires des experts.

Ainsi, il apparaît que minimiser l'erreur quadratique (25) est équivalent à minimiser une mesure de flou des classes ω_i . L'apprentissage conduit alors au système de décision le moins flou possible, au sens de cette mesure de flou.

Fusion graduelle

La méthode proposée permet par ailleurs de traiter les problèmes d'expertises conflictuelles en réalisant automatiquement une fusion graduelle des données de la base d'apprentissage. Supposons en effet que plusieurs experts prononcent des avis différents quant à la classe d'appartenance d'une même observation x . Nous proposons alors d'intégrer tous les avis disponibles à la base d'apprentissage, en présentant séparément chacun des couples :

$$C^q = (x, d^q(x)) \quad (27)$$

$d^q(x)$ désignant le codage net de l'avis de l'expert q pour l'observation x . Les avis des experts seront de la sorte pris en compte dans la proportion de la base d'apprentissage.

Nous proposons d'illustrer cette capacité à fusionner graduellement des avis différents sur un exemple de données synthétiques. Soit $X = [0, 10]^2$ l'univers de référence et soit $\Omega = \{\omega_1, \omega_2\}$ l'ensemble des classes où l'on cherche à affecter tout $x = (x_1, x_2)$ de X . La base d'apprentissage est uniquement constituée d'éléments à composantes entières. On suppose alors qu'à chaque point $x = (x_1, x_2)$, un nombre x_1 d'experts se prononce pour la classe ω_1 alors qu'un nombre x_2 d'experts se prononce pour la classe ω_2 .

Les classes ainsi construites sont alors équiprobables en terme de probabilité *a priori* et les distributions théoriques f_1 et f_2 des probabilités *a posteriori* des classes ω_1 et ω_2 sont connues :

$$f_1(x_1, x_2) = \frac{x_1}{x_1 + x_2} \quad \text{et} \quad f_2(x_1, x_2) = \frac{x_2}{x_1 + x_2} \quad (28)$$

Nous levons l'indétermination à l'origine en posant : $f_1(0, 0) = f_2(0, 0) = 1/2$. L'apprentissage est effectué avec un PMC comprenant 10 neurones sur la couche cachée.

Le résultat obtenu est illustré Figure 5 où l'on trouve à gauche les fonctions d'appartenance théoriques et à droite, les fonctions d'appartenance estimées.

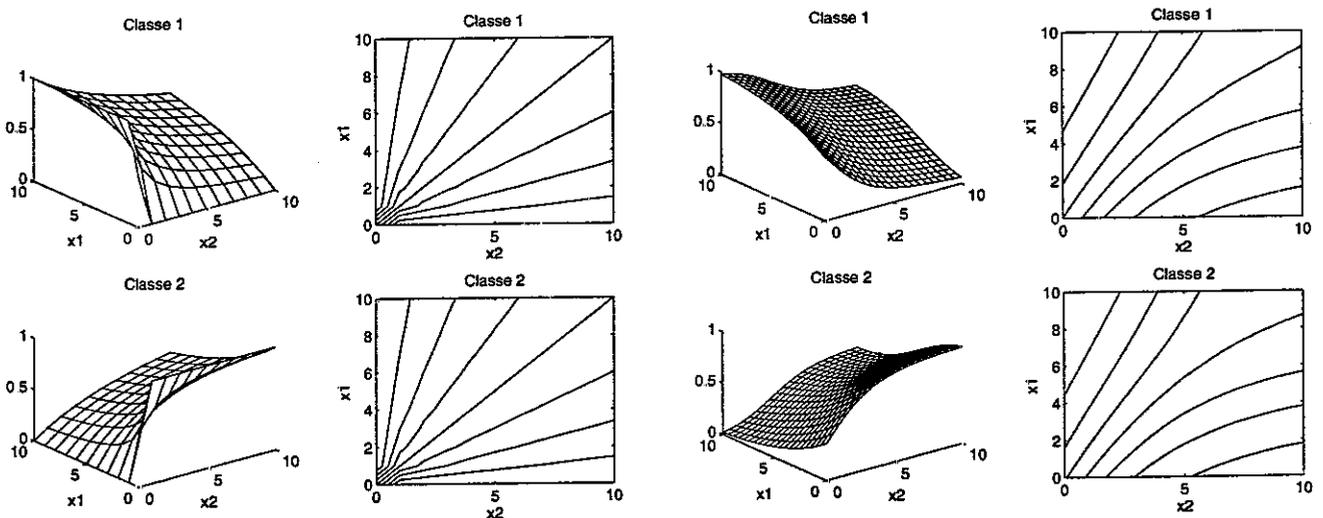


Figure 5

Fusion graduelle des avis d'experts entre conjonction et disjonction

On constate que les degrés d'appartenance sont voisins de 1/2 sur la diagonale, traduisant l'incertitude totale en ces points, puisqu'un nombre égal d'experts s'est prononcé en faveur de chacune des 2 classes. Par ailleurs, l'estimation du degré d'appartenance à ω_i est voisine de 1 lorsque l'observation x fait l'unanimité pour cette classe. C'est le cas pour l'appartenance de (10,0) à la classe ω_1 et de (0,10) à la classe ω_2 .

Un PMC peut donc réaliser une fusion graduelle d'informations, du mode conjonctif où tous les experts sont d'accord, au mode disjonctif où les experts sont en total désaccord.

3.4. Apports de la méthode

La méthode d'estimation de fonctions d'appartenance floues à l'aide de PMC, proposée ci-dessus présente finalement les caractéristiques suivantes :

- utilisation d'un codage simple : codage net de commentaires flous,
- minimisation d'une mesure de flou conduisant à la décision la plus nette,
- fusion graduelle d'informations.

4. Application au confort vibratoire automobile

Nous avons appliqué la méthode exposée précédemment au problème de l'évaluation du confort automobile.

4.1. Contexte et données

L'objectif est de savoir reproduire de façon automatique le diagnostic moyen des experts sur 6 phénomènes (ou encore défauts) en confort automobile :

- les *trépidations* traduisent l'entrée en résonance du bloc moteur ou des roues du véhicules,
- le *tassement* est un défaut correspondant à l'effet cumulé de la suspension sur les butées de chocs et de la fin de course du passager sur son siège.
- le *pompage* est aussi un phénomène basse fréquence, moins prononcé, qui s'amortit sur un nombre de périodes fonction du taux d'amortissement,
- les *mouvements transversaux* correspondent à un roulis violent,
- le *copiage* désigne un faible roulis, en relation avec le profil transversal de route,
- les *mouvements* sur siège traduisent les sensations perçues au niveau du siège par l'occupant.

L'expert confort fait part de ses sensations au cours du temps lors d'essais sur route. La base d'apprentissage est constituée aussi bien des zones temporelles où l'expert s'est prononcé pour l'un des 6 phénomènes décrits ci-dessus, que des zones où il est resté muet. Un couple (*entrée, sortie*) de la base est alors constitué d'un vecteur de 4 paramètres caractéristiques de la zone temporelle et de la classe d'affectation donnée par l'expert. Une zone non signalée sera considérée comme étant étiquetée *sans défaut*. La base de données contient plus de 4000 observations, dont 10% sont gardées pour constituer la base de test.

Une des conditions d'utilisation de la propriété d'estimateur bayésien est d'effectuer l'apprentissage du PMC sur des classes formant une partition. Or, lors d'un essai sur route, si de nombreuses zones sont jugées sans défaut, d'autres au contraire, peuvent présenter plusieurs défauts simultanément (*trépidations* et *pompage* par exemple). Cette relation définissant une partition floue selon [Ruspini 69] n'est donc pas vérifiée lorsque $\Omega = \{\omega_1, \dots, \omega_m\}$ désigne l'ensemble des phénomènes confort. Afin de la satisfaire, nous choisissons de construire un PMC par phénomène, à savoir 6 PMC en tout. Pour chaque phénomène Ω de cardinalité 2, ω_1 désigne alors la classe *défaut* et ω_2 est la classe *non défaut* complémentaire de ω_1 au sens de Zadeh :

$$\mu_{\omega_2}(x) = 1 - \mu_{\omega_1}(x) \quad (29)$$

Si par exemple ω_1 désigne la classe du phénomène de *pompage*, alors ω_2 désigne la classe *non pompage* et la relation (29) signifie qu'une zone temporelle appartient d'autant moins à la classe *pompage* (respectivement *non pompage*) qu'elle appartient d'autant plus à la classe *non pompage* (respectivement *pompage*).

4.2. Mise en œuvre

La méthode d'apprentissage utilisée sur les PMC est la méthode constructive proposée par Lengellé et Denœux [Lengellé 92] mentionnée au paragraphe 3.1, permettant ainsi d'optimiser à la fois la structure des PMC et les poids de leurs connexions. Une validation est en outre mise en œuvre.

Par ailleurs, les effectifs des classes présentent de grosses différences. En effet, une zone temporelle appartient plus souvent à la classe *non défaut* qu'à la classe *défaut*. Pour ne pas favoriser les classes de plus gros effectifs pendant l'apprentissage, le critère (14) fondé sur l'erreur quadratique a été modifié. Il devient alors :

$$J_{\phi_w} = \frac{1}{m} \sum_{i=1}^m J_{\phi_w}^{\omega_i} \quad (30)$$

avec, pour chacune des 2 classes ω_i :

$$J_{\phi_w}^{\omega_i} = \frac{1}{|\omega_i|} \sum_{\substack{k \in \mathcal{F} \\ x \in \omega_i \\ \text{expert}}} (s_k^*(x) - s_k(x))^2 \quad (31)$$

$|\omega_i|$ désignant le nombre d'éléments étiquetés ω_i par l'expert.

4.3. Résultats

Comme il a déjà été suggéré lors du paragraphe consacré à la classification floue, il est difficile de juger du résultat d'une telle classification, en raison de la nature des classes. Notre analyse des résultats s'articule alors autour de la structure des PMC, des valeurs obtenues pour les cardinalités relatives et mesures de flou, et enfin d'une visualisation du comportement de la sortie des PMC au cours du temps lors d'essais sur route.

L'optimisation des 6 PMC par la méthode constructive conduit à des couches cachées de tailles différentes :

- PMC Trépidations : 9 neurones,
- PMC Tassement : 5 neurones,
- PMC Pompage : 7 neurones,
- PMC Mouvements transversaux : 2 neurones,
- PMC Copiage : 8 neurones,
- PMC Mouvements sur siège : 5 neurones.

Le nombre de neurones obtenus sur la couche cachée constitue déjà une bonne indication sur la nature du phénomène à diagnostiquer. Les plus petits PMC sont obtenus pour les *mouvements transversaux*, le *tassement* et les *mouvements sur siège*. Il est intéressant de constater que ces défauts correspondent effectivement, selon les experts en confort automobile, aux phénomènes les plus marqués, donc les plus faciles à reconnaître. Par ailleurs, les *trépidations* constituent un défaut moins ponctuel et moins net. En effet, lorsqu'un véhicule présente des *trépidations*, celles-ci sont en général ressenties pendant toute la durée du parcours, à des degrés plus ou moins élevés. L'expert ne peut alors les mentionner sans arrêt. Cet aspect rend la quantification des *trépidations* plus délicate que celle des autres défauts et explique en partie les 9 neurones de la couche cachée du PMC correspondant.

Le calcul des cardinalités relatives telles qu'elles ont été définies au paragraphe 2.3, relation (10), fournissent aussi une information pertinente sur la nature et sur le degré de flou des classes. Nous rappelons que plus les termes diagonaux ($CR_i(\omega_i)$) sont voisins de 1, plus le phénomène est net et bien reconnu par le PMC. Nous avons regroupé Figure 6, l'ensemble des cardinalités relatives calculées sur les 2 classes *défaut* (classe 1) et *non défaut* (classe 2) de chacun des PMC. Les traits pleins concernent les éléments de la base d'apprentissage alors que les traits pointillés concernent les éléments de la base de test.

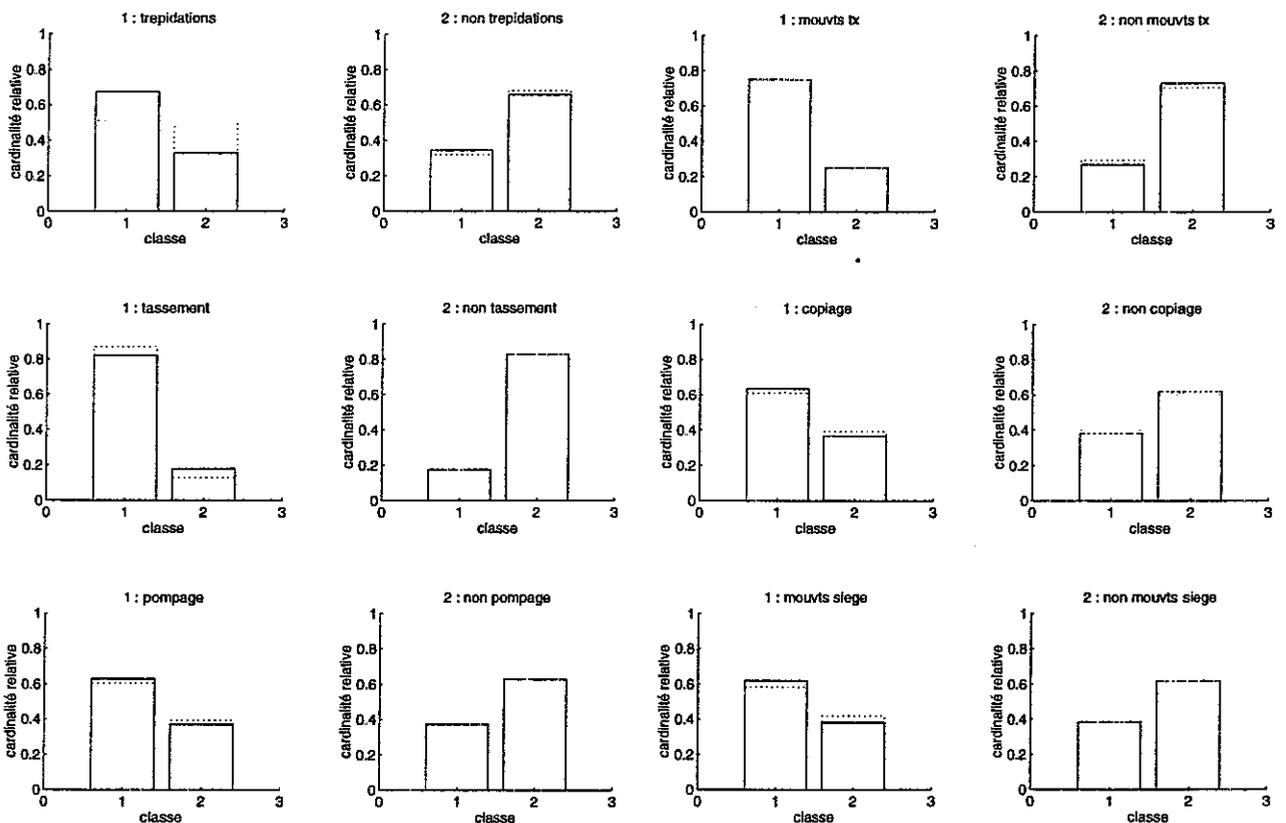


Figure 6
Cardinalités relatives des 6 phénomènes

Pour fixer les idées, détaillons la lecture des diagrammes correspondant au défaut *tassement*. Les éléments de la base d'apprentissage étiquetés *tassement* par les experts recueillent effectivement 80% de la quantité d'appartenance dans cette classe. En revanche, ces éléments appartiennent en moyenne,

avec un degré de 0.2, à la classe *non tassement*. Les courbes en pointillé étant très proches des courbes en trait plein, un commentaire similaire peut être fait sur les éléments de la base de test.

On constate que les résultats obtenus sur la base d'apprentissage et la base de test, en terme de cardinalité relative, sont très voisins, mis à part pour les éléments étiquetés *trépidations* par les experts. Pour cette classe, les éléments de la base de test sont en effet moins bien discriminés que ceux de la base d'apprentissage, sans doute parce que l'ensemble de test n'est pas représentatif de l'ensemble d'apprentissage, malgré les précautions prises sur le découpage de la base de données. Il est par ailleurs rassurant que les éléments non étiquetés *trépidation* par les experts ne soient que faiblement (degré d'appartenance inférieur à 0.4) rattachés à cette classe, et de façon cohérente cette fois entre ensemble d'apprentissage et ensemble test.

D'une façon générale, la Figure 6 confirme que les phénomènes de *mouvements transversaux* et *tassement* sont les moins flous, en accord avec les commentaires les moins ambigus des experts.

La mesure de flou (26) minimisée permet elle aussi de mieux connaître les classes de phénomènes étudiées. La Figure 7 regroupe les mesures de flou calculées à partir des estimations des degrés d'appartenance fournies par les 6 PMC.

La proximité des résultats entre base d'apprentissage et base de test valide, dans une certaine mesure, le découpage de la base de données. En outre, les *mouvements transversaux* se révèlent à nouveau comme les phénomènes les moins flous.

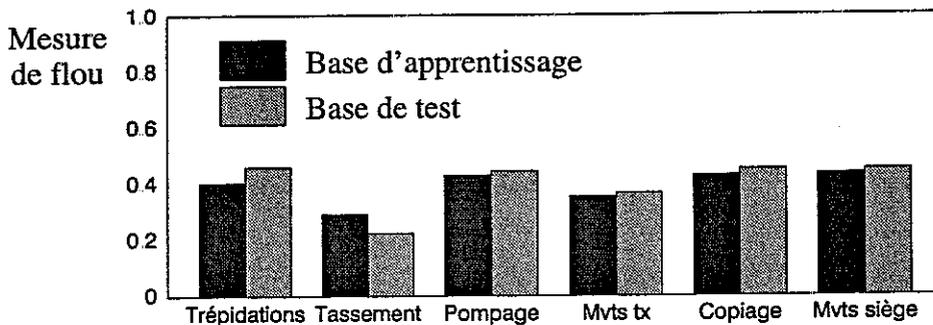


Figure 7
Mesures de flou

Dans ce contexte de qualification de phénomènes à forte composante imprécise et incertaine, la meilleure validation reste sans doute la visualisation au cours du temps du résultat de la mise en œuvre des 6 PMC sur une partie de la base de test. Les Figure 8 et Figure 9 illustrent deux zones temporelles de 5 secondes, extraites de la base de test. Sur ces figures, les signaux représentés à gauche sont les signaux vibratoires recueillis sur le véhicule, à partir desquels les paramètres caractérisant une zone temporelle et alimentant les PMC, sont calculés. A droite sont représentées les évolutions des degrés d'appartenance à chaque phénomène, estimés par les PMC. On y trouve, de haut en bas les *trépidations*, le *tassement*, le *pompage*, les *mouvements transversaux*, le *copiage* et les *mouvements sur siège*. Tout défaut signalé par l'expert est représenté par un trait épais au dessus de la courbe correspondante.

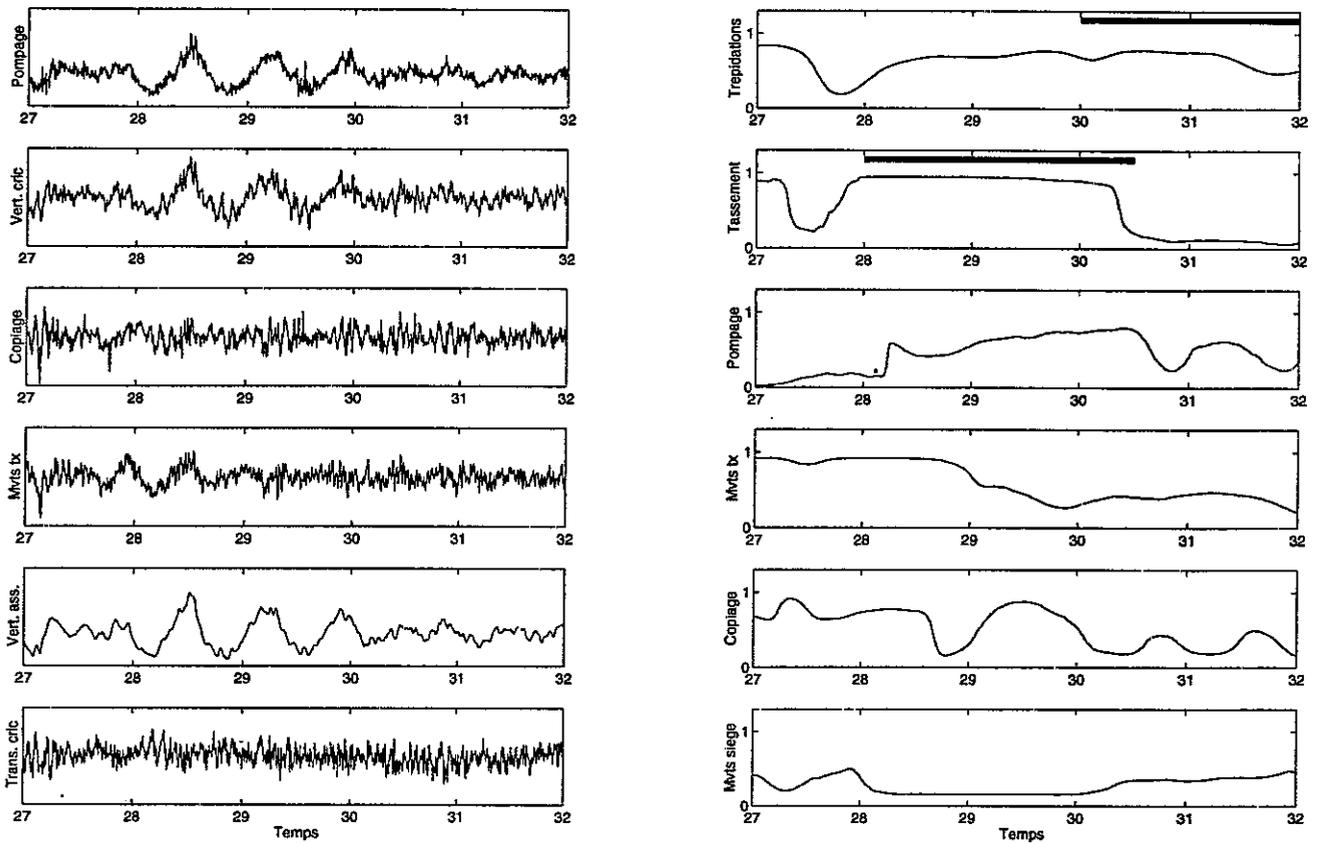


Figure 8
Parcours test : $t_0 = 27s$

La Figure 8 nous montre l'exemple d'une zone où l'expert a signalé successivement un *tassement* et un ensemble de *trepidations*. Cette figure suscite les remarques suivantes :

- Les 2 défauts signalés par l'expert ont été bien reconnus par leurs PMC respectifs.
- La zone de *trepidations* estimée débute bien avant la zone de *trepidations* annoncée par l'expert. En effet, l'expert ne peut signaler qu'un seul phénomène à la fois même s'il en ressent plusieurs simultanément. Il signale alors le plus marqué, c'est-à-dire le tassement dans le cas de la Figure 8. On remarquera en outre que d'autres phénomènes présentent eux aussi un degré d'appartenance estimé élevé, montrant que la zone traversée est particulièrement agitée.
- La nature des courbes d'évolution des estimations de degrés d'appartenance est très variable d'un phénomène à l'autre. On peut notamment constater que la courbe des *trepidations* reste plutôt douce alors que les transitions sont beaucoup plus brutales pour un phénomène comme le *tassement*. Dans ces 2 cas, la nature de la courbe correspond tout à fait à la nature du défaut telle qu'elle est perçue par les experts confort. Ce point très satisfaisant montre que la méthode a permis de retrouver *a posteriori* la nature des phénomènes, sans que celle-ci ne soit explicitement introduite dans le système d'apprentissage.
- Sur la Figure 8, on peut aussi constater que le niveau de *pompage* s'élève progressivement, un peu décalé par rapport au *tassement*. Même si ce *pompage* n'a pas été signalé, sans doute en raison de la violence du *tassement* qui le précède, sa présence n'est pas du tout surprenante car, selon les experts, une zone de *pompage* suit souvent une zone de *tassement*. Cet aspect valide à nouveau, dans une certaine mesure, la méthode proposée.

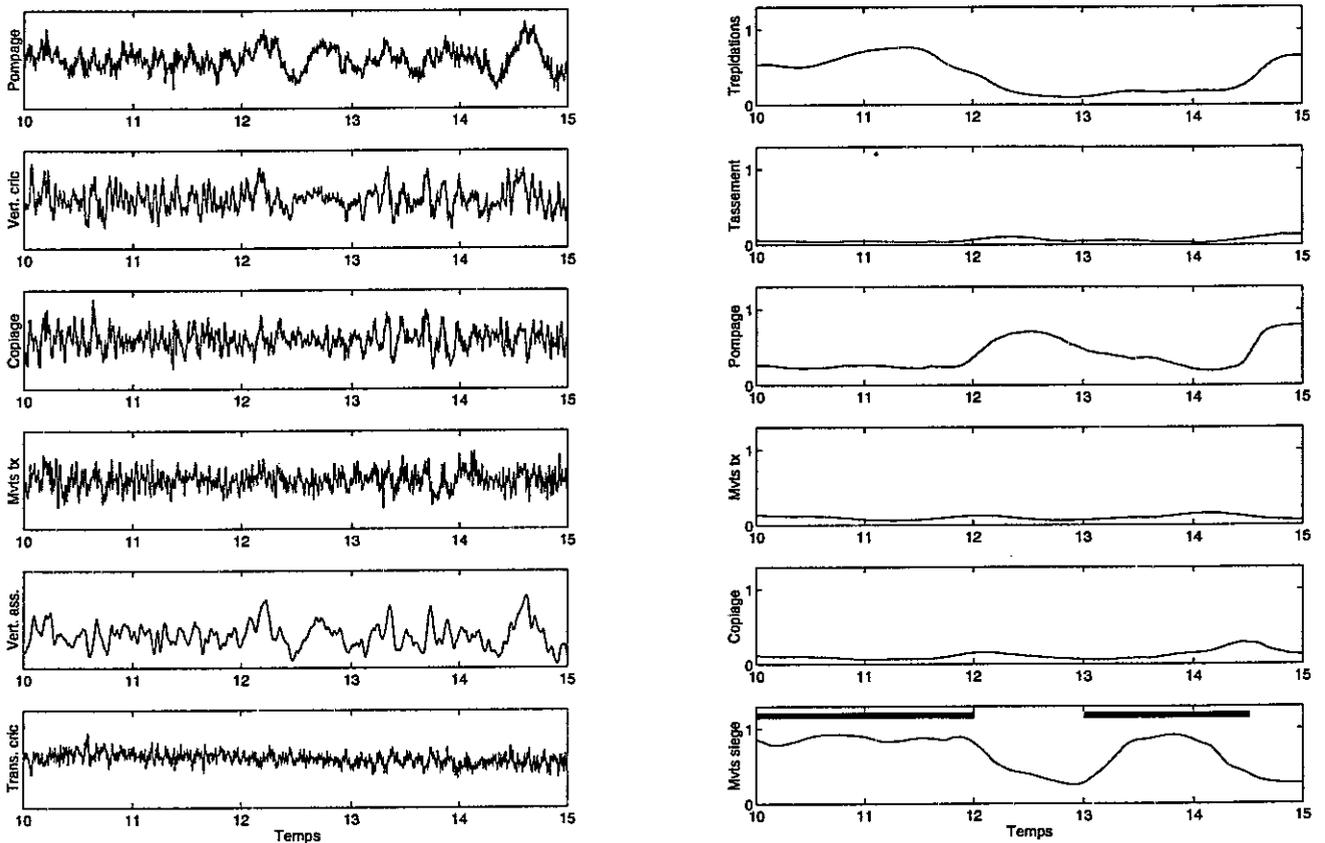


Figure 9
Parcours test : $t_0 = 10s$

La Figure 9 illustre une zone pendant laquelle l'expert a signalé la présence de *mouvements sur sièges*. Là encore les estimations du PMC correspondant à ce phénomène sont bonnes. Par ailleurs, la première zone de *mouvements sur siège* s'accompagne de *trepidations*. Ce point est tout à fait plausible car ces deux phénomènes de fréquences de résonance proches sont jugés assez voisins par les experts. On remarquera en outre que la perception des *mouvements sur siège* a tendance à diminuer dès qu'un phénomène plus basse fréquence (ici le *pompage*) apparaît, conformément à ce que les experts confort ont l'habitude de ressentir.

5. Conclusion

Nous avons montré dans cet article comment exploiter l'expertise humaine pour construire un système d'évaluation automatique de quantités subjectives. La méthode proposée utilise les propriétés des perceptrons multicouche pour quantifier un phénomène sous la forme de l'estimation d'une fonction d'appartenance floue. À partir d'un codage net des expertises humaines, elle consiste à minimiser une mesure de flou et conduit ainsi à la décision la moins floue au sens de cette mesure. En outre, la méthode prend en compte tout type d'information floue (incertitudes probabilistes, résolutionnelles ou floues) contenue implicitement dans ce codage. Nous avons montré et illustré qu'elle permettait également de réaliser une fusion graduelle d'informations, du mode conjonctif où les experts sont d'accord, au mode disjonctif où les experts sont en total désaccord.

Une nouvelle mesure, appelée *cardinalité relative* est proposée pour évaluer le système de décision construit. Elle joue le même rôle que celui des matrices de confusion dans le cadre de la classification nette.

Cet article présente enfin les résultats obtenus sur l'application industrielle relative à l'évaluation du confort dynamique automobile, à partir de données fournies par Renault. Outre la bonne adéquation entre l'estimation fournie par le système et celle de l'expert, il est remarquable de retrouver une certaine connaissance des experts dans le comportement du système automatique d'évaluation. C'est le cas par exemple pour la nature des phénomènes ou pour leurs relations de cause à effet. Ces observations constituent ainsi une validation *a posteriori* de la méthode, dont les perspectives peuvent être étendues à toute application où l'expertise humaine, et d'une manière générale la subjectivité, sont impliquées.

Références

- [Bottou 91] : L. Bottou, *Une approche théorique de l'apprentissage connexionniste ; applications à la reconnaissance de la parole*. Thèse de doctorat de l'université de Paris Sud, 1991.
- [Bourlard 88] : H. Bourlard and Y. Kamp, "Auto-association by multilayer perceptrons and singular value decomposition," *Biological cybernetics*, vol. 59, pp. 291–294, 1988.
- [Bouchon-Meunier 93] : B. Bouchon-Meunier, *La logique floue*. Paris : PUF, 1993.
- [Chabanon 92] : C. Chabanon, Y. Lechevallier and S. Millemann, "An efficient neural network by a classification tree," *Proceedings COMPSTAT'92*, vol. 1, pp. 227–232, Neuchâtel, 1992.
- [Ciampi 95] : A. Ciampi et Y. Lechevallier, "Réseaux de neurones et modèles statistiques," *La revue de Modulad*, n°15, pp. 27–46, 1995.
- [Cybenko 89] : G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of control, signals and systems*, vol. 2, n°4, pp. 303–314, 1989.
- [Dubois 80] : D. Dubois and H. Prade, *Fuzzy sets and systems : theory and applications*. London : Academic Press, 1980.
- [Fahlman 90] : S. E. Fahlman and C. Lebiere, "The cascade-correlation learning architecture," *Advances in neural information processing systems II*, ed. D. S. Touretsky, pp. 524–532, San Mateo, 1990.
- [Gallinari 91] : P. Gallinari, S. Thiria, F. Badran and F. Fogelman Soulié, "On the relations between discriminant analysis and multilayer perceptrons," *Neural Networks*, vol. 4, pp. 349–360, 1991.
- [Hertz 91] : J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the theory of neural computation*. Santa Fe : Addison Wesley, 1991.
- [Hornik 89] : K. Hornik, M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, 1989.
- [Judd 90] : J. S. Judd, *Neural network design and the complexity of learning*. MIT Press, Cambridge, MA, 1990.
- [Kaufmann 73] : A. Kaufmann, *Introduction à la théorie des sous-ensembles flous*. Paris : Masson, 1973.

- [Kosko 92] : B. Kosko, *Neural networks and fuzzy systems*. Prentice-Hall International, 1992.
- [Le Cun 90] : Y. Le Cun, J. S. Denker and S. A. Solla, "Optimal brain damage," In D. Touretzky, editor, *Advances in neural information processing systems 2*, pp. 598-605, 1990.
- [Lengellé 92] : R. Lengellé and T. Dencœux, "Optimizing multi-layer network layer by layer without backpropagation," *ICANN'92*, 1992.
- [Millemann 91] : S. Millemann, *Construction et initialisation d'un réseau de neurones à partir d'un arbre de segmentation*, Mémoire de DEA Contrôle des systèmes UTC/INRIA, 1991.
- [Millemann 95] : S. Millemann and R. Lengellé, "Fuzzy supervised membership estimation using a multilayer perceptron," *Proc. EUFIT'95*, vol. I, pp. 538-542, Aachen, 1995.
- [Millemann 96] : S. Millemann, *Estimation connexionniste et fusion possibiliste de fonctions d'appartenance floue. Application à l'évaluation du confort automobile*. Thèse, Université de Technologie de Compiègne, Compiègne, 1996.
- [Mtimet 94] : R. Mtimet, *Techniques de régularisation appliquées aux réseaux de neurones*, Mémoire de DEA INSA Toulouse, 1994.
- [Nowakowska 77] : M. Nowakowska, "Fuzzy concepts in the social sciences," *Behav. Sci.*, vol. 22, pp. 107-115, 1977.
- [Pal 94] : N. R. Pal and J. C. Bezdek, "Measuring fuzzy uncertainty," *IEEE Transactions on Fuzzy Systems*, vol. 2, pp. 107-118, 1994.
- [Richard 91] : M. D. Richard, and R. P. Lippmann, "Neural network classifiers estimate bayesian a posteriori probabilities," *Neural computation*, vol. 3, pp. 461-483, 1991.
- [Ruck 90] : D. W. Ruck, S. K. Rogers, M. Kabrisky, M. E. Oxley, and B. W. Suter, "The multilayer perceptron as an approximation to a Bayes optimal discriminant function," *IEEE Trans. Neural networks*, vol. 1, n°4, pp. 296-298, 1990.
- [Rumelhart 86] : D. E. Rumelhart and J. L. McClelland, *Parallel distributed processing*. Cambridge, MA : MIT press, 1986.
- [Ruspini 69] : E. H. Ruspini, "A new approach to clustering," *Inf. Control*, vol. 15, pp. 22-32, 1969.
- [Sethi 90] : I. K. Sethi, "Entropy nets : from decision trees to neural networks," *Proceedings of IEEE*, vol. 78, 1990.
- [Shafer 76] : G. Shafer, *A mathematical theory of evidence*. Princeton University Press, Princeton, New Jersey, 1976.
- [Webb 90] : A. R. Webb and D. Lowe, "The optimised internal representation of multilayer classifier networks performs non linear discriminant analysis," *Neural Networks*, vol. 3, pp. 367-375, 1990.
- [Wonnacott 91] : T. Wonnacott and R. Wonnacott, *Statistique*, Economica, 1991.
- [Zadeh 65] : L. A. Zadeh, "Fuzzy sets," *Inform. control*, vol. 8, pp. 338-353, 1965.
- [Zadeh 84] : L. A. Zadeh, *Fuzzy sets and commonsense knowledge*. Berkeley, CA: University of California Berkeley, Cognitive science report n°21, 1984.