

Pour action :

M. Albertelli	GIST	FR TCR GRA 0 90
M. Sidorkiewicz	64240	FR TCR GRA 0 90
M. Visconti	64240	FR TCR GRA 0 90
Mlle. Huszar	GIST	FR TCR GRA 0 90
M. Daumard	64240	FR TCR GRA 0 90
Mlle. Ansaldi	64240	FR TCR GRA 0 90

Pour info :

M. Le Douaron	64240	FR TCR GRA 0 90
---------------	-------	-----------------

Pour archivage :

Mme. Zeender	64240	FR TCR GRA 0 90
--------------	-------	-----------------



RENAULT

DR / Groupe optimisation

Yann COLLETTE - Service 64240

API : FR TCR GRA 0 90

Tél. : 01 76 85 81 93

Fax : 01 76 85 77 13

E-Mail : yann.collette@renault.com

Guyancourt, le 1 avril 2005

Note technique n° 84/2005/64240

Thème : Le Gamma Test

Objectif : Utilisation d'une méthode permettant d'estimer l'erreur de modélisation d'une relation entrée / sortie

Attente : Votre information et prise en compte

Résumé - Synthèse :

Lorsque l'on cherche à modéliser une relation entrée / sortie, on est confronté à plusieurs choix cruciaux :

- quelles entrées sélectionner pour obtenir un bon modèle ;
- quelle structure donner à un réseau de neurones ;
- etc.

La méthode du Gamma Test permet de répondre à ces questions sans que l'on ait à produire de réseau de neurones pour tester une configuration d'entrées. Cette méthode est simple à implémenter et produit des estimations du bruit de modélisation fiables.

Le Gamma Test

Y. Collette

Table des matières

1	Introduction	1
2	Algorithme	2
3	La représentation des informations du Gamma Test	4
3.1	Introduction	4
3.2	Le graphe de proximité	4
3.3	Le M-Test	5
4	Utilisation	6
4.1	Sélection des entrées	6
4.1.1	La sélection exhaustive d'entrées	6
4.1.2	La sélection gloutonne "forward"	7
4.1.3	La sélection gloutonne "backward"	8
4.1.4	Conclusion	9
4.2	Mesure de l'information résiduelle	9
4.3	Mesure du modèle sous-jacent	10
4.4	Estimation de la structure minimale d'une fonction paramétrique	11
4.5	Analyse de sensibilité	13
5	Conclusion	14

1 Introduction

Lorsque l'on cherche à modéliser un système en utilisant des fonctions paramétriques (comme les réseaux de neurones, les polynômes, etc.), on n'obtient jamais de modèle parfait. Ceci est dû principalement à deux choses :

- une méconnaissance du système (on ne maîtrise pas toutes les interactions entre les constituants du système) ;
- la présence de bruit dans les mesures des sorties (et des entrées).

La sortie $u_s(t)$ que l'on cherchera à modéliser se décomposera donc de la façon suivante :

$$u_s^{\text{mesuré}}(t) = u_s^{\text{réel}}(t) + b(t) \quad (1)$$

Il est important de pouvoir connaître a priori le niveau de bruit dans les mesures car la connaissance de cette information va nous permettre de savoir quand arrêter le processus d'apprentissage de notre fonction paramétrique.

En effet, si on s'arrête trop tard, la fonction paramétrique va "apprendre" une portion du bruit et on verra, à ce moment, apparaître un phénomène que l'on appelle "surapprentissage". La fonction paramétrique a trop bien appris les mesures et reproduit le signal d'erreur contenu dans les mesures. Ce qui signifie que si on présente à la fonction paramétrique des données qu'elle n'aura pas vu lors de la phase d'apprentissage, mais pour lesquelles la fonction paramétrique sera capable de générer une estimation de la mesures, la fonction paramétrique va commettre une erreur. Elle

va “ajouter” du bruit à l’estimation et ainsi générer une mauvaise estimation (c’est une description imagée du phénomène de surapprentissage).

Etre en mesure de pouvoir quantifier cette part de bruit dans les mesures est donc un atout vital qui peut nous permettre d’obtenir de meilleurs modèles de nos systèmes physiques.

Il peut sembler surprenant de pouvoir quantifier a priori le niveau d’un bruit de mesure sans avoir aucune information à son sujet. Cependant, un tel outil existe : c’est le Gamma Test (voir [Durrant 01] ou [Tsui 99]).

2 Algorithme

On a un jeu de données représenté de la façon suivante :

$$((x_1, \dots, x_m), y) = (\vec{x}, y) \quad (2)$$

m correspond à la dimension du vecteur \vec{x} .

On suppose que ce jeu de données peut être modélisé par une fonction paramétrique lisse f (fonction dont les dérivées partielles premières et secondes sont bornées - il n’y a pas de discontinuité dans la valeur de y quand on applique une petite variation sur \vec{x}). Cette hypothèse d’existence d’une borne sur les dérivées premières et secondes s’écrit de la façon suivante :

$$\nabla f(\vec{x}) = \sum_{i=1}^m \frac{\partial f(\vec{x})}{\partial x_i} \cdot \vec{e}_i \quad \text{et} \quad Hf(\vec{x}) = \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 f(\vec{x})}{\partial x_i \cdot \partial x_j} \cdot \vec{e}_j$$

L’hypothèse “ f est une fonction lisse” s’écrit :

$$\exists b_1 > 0, b_2 > 0 \text{ telles que } \forall \vec{x} \in \mathbb{R}^m \text{ on ait } |\nabla f(\vec{x})| \leq b_1 \text{ et } |Hf(\vec{x})| < b_2 \quad (3)$$

La relation existant entre tous ces éléments est la suivante :

$$y = f(\vec{x}) + r \quad (4)$$

où r correspond à un bruit (qui peut être soit un bruit de mesure sur y , soit un bruit de modélisation, soit les deux).

Si on note $x(N(i, p))$ le $p^{\text{ième}}$ plus proche voisin du $i^{\text{ième}}$ point de la base de données, le calcul de l’estimation de la variance du bruit passe par le calcul des expressions suivantes :

$$\gamma_n(p) = \frac{1}{2 \cdot n} \cdot \sum_{i=1}^n |y_{N(i,p)} - y_i|^2 \quad (5)$$

$$\delta_n(p) = \frac{1}{n} \cdot \sum_{i=1}^n |\vec{x}(N(i, p)) - \vec{x}(i)|^2 \quad (6)$$

L’expression $|\vec{x}(i) - \vec{x}(j)|$ correspond à la distance entre le point $\vec{x}(i)$ et le point $\vec{x}(j)$:

$$|\vec{x}(i) - \vec{x}(j)| = \sum_{k=1}^m |\vec{x}_k(i) - \vec{x}_k(j)| \quad (7)$$

Les voisins d’un points $\vec{x}(i)$ sont extraits au moyen d’une structure du type KD-Tree. Cette structure correspond à un arbre dans lequel on va stocker tout le jeu de données et, dans cet arbre, l’extraction d’un voisin d’un point donné se fait en $o(M \cdot \log(M))$ avec M correspondant au nombre de points à extraire.

La procédure à suivre pour le calcul du Gamma Test est alors la suivante :

- On calcule les valeurs du Gamma Test pour p variant de 1 à p_{max} (avec, typiquement, $p_{max} \simeq 20$).
- On reporte ces valeurs sur un graphique $\gamma_n = g(\delta_n)$.
- On calcule la droite de régression $\gamma_n = A \cdot \delta_n + B$.

La variance du bruit r s'identifie au paramètre B de la droite de régression. Le coefficient A correspond à l'expression suivante :

$$A(n, p) = \frac{\xi_\phi \left(((\vec{x}(N(i, p)) - \vec{x}(i)) \cdot \nabla f(\vec{x}(i)))^2 \right)}{2 \cdot \xi_\phi(|\vec{x}(N(i, p)) - \vec{x}(i)|)} \quad (8)$$

avec :

- ϕ la distribution de probabilité de sélection des points dans le voisinage du point $\vec{x}(i)$;
- n le nombre de points sélectionnés suivant la distribution ϕ ;
- p la “distance” par rapport au point courant $\vec{x}(i)$.

On note $\xi_\phi(r)$ l'espérance de r suivant la distribution de probabilité ϕ .

Exemple : si r est un bruit gaussien centré de variance σ . Alors la distribution ϕ est une distribution gaussienne de moyenne nulle et de variance ϕ . On en déduit donc que $\xi_\phi(r) = 0$.

Alors, si on note θ l'angle que forme le vecteur $\vec{x}(N(i, p)) - \vec{x}(i)$ (ce vecteur est normalement contenu dans le plan tangent au point $\vec{x}(i)$ si $p \rightarrow 0$) avec le vecteur $\nabla f(\vec{x}(i))$ (qui est normalement perpendiculaire au plan tangent au point $\vec{x}(i)$ si $p \rightarrow 0$) alors on peut réécrire l'expression de $A(n, p)$ de la façon suivante :

$$A(n, p) = \frac{1}{2} \cdot \langle |\nabla f|^2 \rangle \cdot \langle \cos^2 \theta \rangle \quad (9)$$

Le coefficient $A(n, p)$ correspond donc, à un facteur près, à la moyenne de la norme du carré du gradient. Cette expression a une certaine importance car elle nous servira par la suite à déterminer la structure minimale d'un réseau de neurones.

Dans la pratique, si les points \vec{x} sont distribués uniformément dans le domaine de définition de f , la valeur moyenne du carré du cosinus vaut 1. On obtient alors une proportionnalité directe entre le coefficient $A(n, p)$ et la moyenne de la norme du carré du gradient.

A partir de l'estimation de la variance du bruit, il est possible d'obtenir diverses informations :

- le ratio des variances :

$$VR = \frac{Var(r)}{Var(y)} \quad (10)$$

Le ratio des variances (VR Variance Ratio) correspond à une valeur normalisée du bruit de mesure.

- le R_{GT}^2 :

$$R_{GT}^2 = 1 - VR^2 = 1 - \frac{Var(r)^2}{Var(y)^2} \quad (11)$$

Le R_{GT}^2 a la même signification que le R^2 classique, mise à part qu'il est plus adapté aux modèles non-linéaires.

On trouve aussi une autre formule pour le calcul de $\gamma(p)$ et de $\delta(p)$ (voir [Evans 02]) :

$$\delta(p) = \frac{1}{p} \cdot \sum_{i=1}^p \frac{1}{M} \cdot \sum_{j=1}^M |\vec{x}(N(j, i)) - \vec{x}(j)| \quad (12)$$

$$\gamma(p) = \frac{1}{p} \cdot \sum_{i=1}^p \frac{1}{2 \cdot M} \cdot \sum_{j=1}^M |y(N(j, i)) - y(j)| \quad (13)$$

L'idée portée par ces formules est de réduire l'importance des voisins distants.
 Dans nos différents tests, nous n'utiliserons que les formules 5 et 6.

3 La représentation des informations du Gamma Test

3.1 Introduction

Une fois le Gamma Test appliqué à un jeu de données, il est possible de représenter graphiquement les informations fournis par celui-ci.

Par exemple, le paramètre p_{max} a une grande importance dans la qualité de l'estimation fournie par le Gamma Test. Il est possible de voir si la valeur retenue pour p_{max} est raisonnable ou non en traçant le M-Test. Le M-Test est un graphique sur lequel on représente l'estimation de $Var(r)$ pour différentes valeurs de p . Ce graphique peut être calculé très rapidement pour p variant de 1 à p_{max} car toutes les informations nécessaires au calcul de ces estimations sont utilisées dans le calcul de l'estimation de $Var(r)$ pour $p = p_{max}$.

Il peut aussi être intéressant de tracer les points $|\vec{x}(N(i, p)) - \vec{x}(i)|, \frac{1}{2} \cdot |y(N(i, p)) - y(i)|$ sur un graphe et de tracer aussi les points ayant servi au calcul de la droite de régression $(\delta_n(p), \gamma_n(p))$ pour p variant de 1 à p_{max} . Ce graphique est appelé "scatterplot" ou graphe de proximité. Ce graphique représente la corrélation spatiale entre différents points. Si δ est petit, cela veut dire que les points considérés sont proches les uns des autres. Donc, si la sortie n'est pas bruitée, deux points proches dans l'espace des \vec{x} doivent avoir un γ petit. En effet, la variation de valeur de la fonction objectif associée à ces deux points doit être, elle aussi, petite. Par contre, si la sortie est bruitée, deux points proches dans l'espace des \vec{x} n'ont pas forcément un γ petit. Le bruit peut "créer" des variations arbitraires sur la valeur de la fonction objectif.

3.2 Le graphe de proximité

Nous allons appliquer le Gamma Test à un cas simple : une période de sinusoïde. Dans un premier temps, nous appliquerons le Gamma Test à une sinusoïde non bruitée (voir sur la figure 1-a) puis, dans un second temps, à une sinusoïde bruitée (voir sur la figure 1-b).

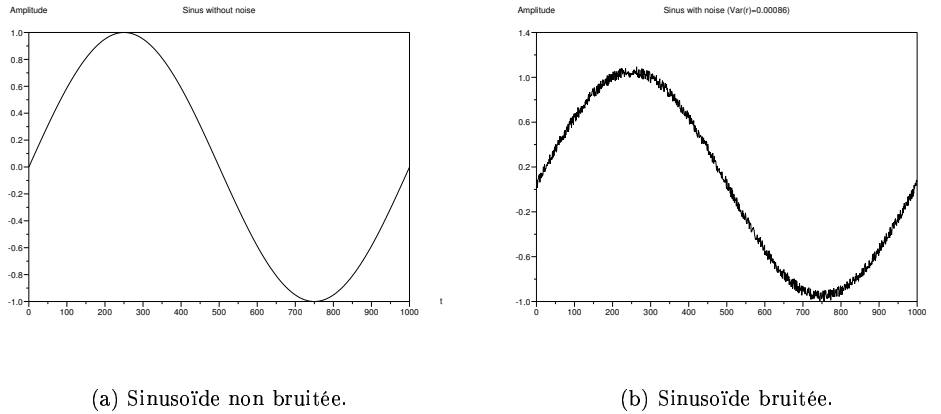


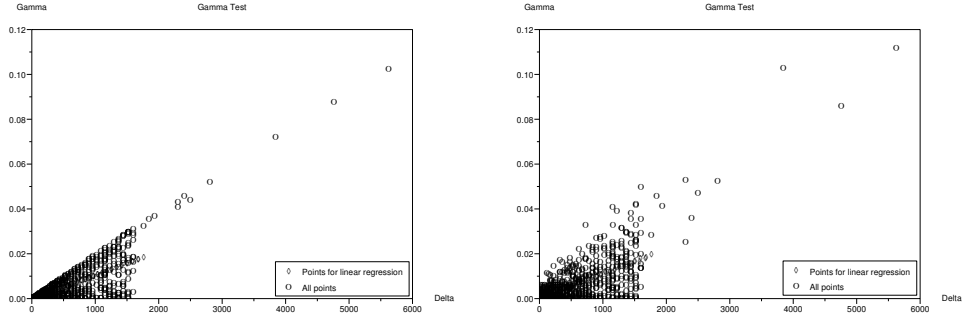
FIG. 1 – L'allure temporelle de la sinusoïde non bruitée et de la sinusoïde bruitée.

Nous allons maintenant tracer sur la figure 2 :

– les points :

$$\left(|\vec{x}(N(i, p)) - \vec{x}(i)|, \frac{1}{2} \cdot |y(N(i, p)) - y(i)| \right) \quad (14)$$

- les points ayant servi au calcul de la droite de corrélation.



(a) Graphe de proximité de la sinusoïde non bruitée.

(b) Graphe de proximité de la sinusoïde bruitée.

FIG. 2 – Les graphiques de proximité correspondant à la sinusoïde bruitée et à la sinusoïde non bruitée.

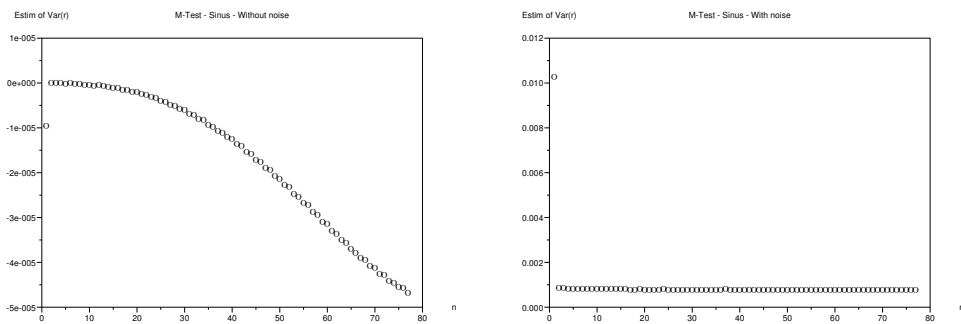
L'estimation de la variance du bruit fournie par le Gamma Test est la suivante :

- pour la sinusoïde non bruitée : $Var(r) = -8.06 \cdot 10^{-7}$
- pour la sinusoïde bruitée : $Var(r) = 0.00086$

La variance du bruit que nous avons ajouté à la sinusoïde valait : $Var(r) = 0.00087$.

3.3 Le M-Test

Nous allons maintenant représenter le M-Test pour la sinusoïde non bruitée (voir sur la figure 3-a) ainsi que la sinusoïde bruitée (voir sur la figure 3-b) pour p variant de 2 à 80.



(a) Graphique du M-Test pour la sinusoïde non bruitée.

(b) Graphique du M-Test pour la sinusoïde bruitée.

FIG. 3 – Le M-Test pour la sinusoïde bruitée et la sinusoïde non bruitée.

On constate ici que la valeur traditionnelle de $p_{max} = 20$ n'est pas tout à fait adaptée pour que l'application du Gamma Test fournisse une bonne estimation de la variance du bruit. Cependant,

l'ordre de grandeur de l'estimation de la variance du bruit reste raisonnable.

4 Utilisation

4.1 Sélection des entrées

Nous allons maintenant utiliser le Gamma Test pour effectuer la sélection d'entrées pour une fonction paramétrique (un réseau de neurones par exemple). Nous allons travailler sur un exemple simple :

$$f(x, y) = x^2 + y^2 \quad (15)$$

Nous allons générer un fichier de données qui sera constitué des valeurs de $f(x, y)$ associées à des valeurs de x et y tirées aléatoirement et uniformément dans l'intervalle $[0, 1]$ (voir sur la figure 4).

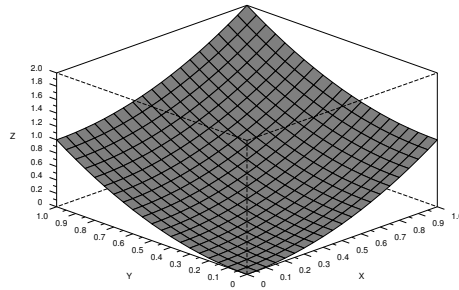


FIG. 4 – Représentation du jeu de données utilisé.

Nous allons générer un second fichier de données dans lequel nous ajouterons une colonne de valeurs t pour t variant de 1 à 1000. Ce fichier supplémentaire nous permettra de vérifier que le Gamma Test permet bien de ne pas retenir les entrées qui ne servent à rien.

4.1.1 La sélection exhaustive d'entrées

La méthode exhaustive de sélection des entrées est l'approche la plus simple que l'on puisse envisager (et certainement la plus naïve). Elle consiste à ajouter et retirer toutes les entrées potentielles à partir d'un modèle à une entrée, de façon à tester toutes les combinaisons possibles. Par exemple, si on a 10 entrées potentielles, l'application de la méthode de sélection exhaustive des entrées va générer le test de $2^{10} = 1024$ combinaisons. Cela peut représenter un temps de calcul important !

Les résultats de l'application de la sélection des entrées par Gamma Test donne :

– pour le fichier de données à deux entrées :

La première colonne du jeu de données correspond à la variable x . La seconde colonne du jeu de données correspond à la variable y .

Masque	$Var(r)$	R_{GT}^2	Comb. retenue à cette étape	Comb. retenue finale
10	0.08819	0.5058		
01	0.0837	0.5309		
11	0.00010	0.9994	×	×

On constate que, d'après le Gamma Test, la meilleure combinaison à retenir correspond à la sélection des entrées x et y .

- pour le fichier de données à trois entrées :

La première colonne du jeu de données correspond à la variable x . La seconde colonne du jeu de données correspond à la variable y . La troisième colonne du jeu de données correspond à la colonne qui n'est pas utilisée dans le modèle : t .

Masque	$Var(r)$	R_{GT}^2	Comb. retenue à cette étape	Comb. retenue finale
100	0.494	0.5058		
010	0.469	0.5309		
110	0.00058	0.9994	×	×
001	0.997	0.003		
101	1.007	-0.007		
011	1.007	-0.007		
111	1.01	-0.01		

On constate que le Gamma test a bien retenu la combinaison 110 qui correspond à la sélection des entrées x et y . Toutes les combinaisons qui incluent la colonne t augmentent fortement le niveau du bruit (l'estimation de la variance $Var(r)$ augmente fortement pour toutes les combinaisons incluant la colonne t).

4.1.2 La sélection gloutonne “forward”

Le principe de cette méthode de sélection est le suivant :

1. On calcule le Gamma Test en sélectionnant les entrées potentielles qui n'ont pas encore été retenues les unes après les autres (en n'oubliant pas d'ajouter la colonne des valeurs de fonction objectif).

On teste ainsi toutes les combinaisons possibles composées d'une entrée supplémentaire.

2. On retient l'entrée qui permet de réduire le plus la valeur de l'estimation de la variance du bruit $Var(r)$.

Une fois cette entrée retenue, on ne remet plus en cause la sélection de cette entrée.

3. S'il reste des entrées à ajouter, on recommence à l'étape 1.

Sinon, on arrête. On retourne alors la meilleure combinaison trouvée.

L'application de ce principe de sélection à l'exemple simple de la section précédente nous donne les résultats suivants :

- Exemple du jeu de données sans l'entrée supplémentaire :
 - Etape 1 : sélection de la première entrée

Masque	$Var(r)$	R_{GT}^2	Comb. retenue à cette étape	Comb. retenue finale
10	0.0882	0.5054		
01	0.0841	0.5286	×	

- Etape 2 : sélection de la seconde entrée

Masque	$Var(r)$	R_{GT}^2	Comb. retenue à cette étape	Comb. retenue finale
11	$1.11 \cdot 10^{-5}$	0.99993	×	

On constate que cette méthode de sélection gloutonne “forward” a permis de sélectionner la bonne combinaison d'entrées.

- Exemple du jeu de données avec l'entrée supplémentaire :
 - Etape 1 : sélection de la première entrée

Masque	$Var(r)$	R_{GT}^2	Comb. retenue à cette étape	Comb. retenue finale
100	0.0882	0.5054		
010	0.0841	0.5286	×	
001	0.1805	-0.0115		

– Etape 2 : sélection de la seconde entrée

Masque	$Var(r)$	R_{GT}^2	Comb. retenue à cette étape	Comb. retenue finale
110	$1.11 \cdot 10^{-5}$	0.99994	×	×
011	0.1844	-0.033		

– Etape 2 : sélection de la seconde entrée

Masque	$Var(r)$	R_{GT}^2	Comb. retenue à cette étape	Comb. retenue finale
111	0.1857	-0.041	×	

On constate que cette méthode de sélection gloutonne “forward” a permis de sélectionner la bonne combinaison d’entrées (masque correspondant à la meilleure sélection d’entrées : 110).

4.1.3 La sélection gloutonne “backward”

Le principe de cette méthode de sélection est le suivant :

1. On calcule le Gamma Test en supprimant les entrées potentielles qui n’ont pas encore été supprimées les unes après les autres (en n’oubliant pas d’ajouter la colonne des valeurs de fonction objectif).

On teste ainsi toutes les combinaisons possibles composées d’une entrée supplémentaire.

2. On supprime l’entrée qui permet de réduire le plus la valeur de l’estimation de la variance du bruit $Var(r)$.

Une fois cette entrée supprimée, on ne remet plus en cause la suppression de cette entrée.

3. S’il reste plus d’une entrée à supprimer, on recommence à l’étape 1.

Sinon, on s’arrête. On retourne alors la meilleure combinaison trouvée.

L’application de ce principe de sélection à l’exemple simple de la section précédente nous donne les résultats suivants :

– Exemple du jeu de données sans l’entrée supplémentaire :

– Etape 1 : calcul de la variance pour la combinaison de départ

Masque	$Var(r)$	R_{GT}^2	Comb. retenue à cette étape	Comb. retenue finale
11	$1.11 \cdot 10^{-5}$	0.99993	×	×

– Etape 2 : suppression de la première entrée

Masque	$Var(r)$	R_{GT}^2	Comb. retenue à cette étape	Comb. retenue finale
10	0.0882	0.5054		
01	0.0841	0.5286	×	

On constate que cette méthode de sélection gloutonne “backward” a permis, tout comme la méthode de sélection “forward”, de sélectionner la bonne combinaison d’entrées.

– Exemple du jeu de données avec l’entrée supplémentaire :

– Etape 1 : calcul de la variance pour la combinaison de départ

Masque	$Var(r)$	R_{GT}^2	Comb. retenue à cette étape	Comb. retenue finale
111	0.1857	-0.041	×	

– Etape 2 : suppression de la première entrée

Masque	$Var(r)$	R_{GT}^2	Comb. retenue à cette étape	Comb. retenue finale
110	$1.11 \cdot 10^{-5}$	0.99994	×	×
101	0.1839	-0.031		
011	0.1844	-0.033		

– Etape 3 : suppression de la seconde entrée

Masque	$Var(r)$	R_{GT}^2	Comb. retenue à cette étape	Comb. retenue finale
100	0.0882	0.5054		
010	0.0841	0.5286	×	

On constate que cette méthode de sélection gloutonne “backward” a permis de sélectionner la bonne combinaison d’entrées (masque correspondant à la meilleure sélection d’entrées : 110).

4.1.4 Conclusion

Comme on peut le constater, le Gamma Test permet de sélectionner les entrées les plus pertinentes dans le but d’effectuer la modélisation d’un phénomène.

Plusieurs méthodes permettent d’effectuer cette sélection :

- la méthode de sélection exhaustive : cette méthode est la plus rigoureuse car on teste toutes les combinaisons possibles d’entrées. Si on a beaucoup d’entrées potentielles à tester, le nombre de tests à effectuer peut faire exploser le temps de calcul.
- la méthode de sélection gloutonne : cette méthode permet de ne tester qu’une partie des combinaisons possibles. Cependant, de part son principe de fonctionnement, il est possible de “manquer” la bonne combinaison. Par exemple, si la valeur de p_{max} est trop petite, il est possible que l’erreur sur l’estimation du bruit nous fasse sélectionner une entrée à la place d’une autre.

Cette erreur sur l’estimation de la variance du bruit peut avoir un impact sur la sélection d’entrées par la méthode exhaustive. Cependant, la méthode de sélection exhaustive est moins sensible à ce genre d’erreur que la méthode de sélection gloutonne.

Le principal avantage de la méthode de sélection gloutonne vient du fait que l’on ne teste pas toutes les combinaisons possibles d’entrées. Cette méthode est donc plus rapide à appliquer.

4.2 Mesure de l’information résiduelle

Nous allons maintenant reprendre l’exemple simple de la sinusoïde bruitée. Nous allons volontairement “mal” modéliser cette sinusoïde par une autre sinusoïde de plus faible amplitude :

$$y = A \cdot \sin(2\pi \cdot t)$$

Nous allons utiliser le Gamma Test pour calculer le R_{GT}^2 du résidu pour voir s’il reste de l’information dans ce résidu. Si le R_{GT}^2 du résidu est proche de 1, cela veut dire qu’il reste une quantité d’information non négligeable et que notre apprentissage de la sinusoïde est mauvais. Il nous faut impérativement modifier l’amplitude de notre sinusoïde pour récupérer de l’information.

Nous allons donc calculer le R_{GT}^2 du résidu (différence entre l’estimation fournie par le modèle et la mesure) pour différentes valeurs de A s’approchant de la bonne valeur de A : 1.

Amplitude	$Var(r)$	R_{GT}^2
0.8	0.00086	0.9589
0.9	0.00086	0.8537
0.95	0.00086	0.5944
0.99	0.00086	0.0663

On constate que le Gamma Test, par l'intermédiaire du R_{GT}^2 , nous indique qu'il reste de l'information dans le résidu. Par exemple, lorsque l'on choisit une amplitude de 0.8 pour notre modèle, la Gamma Test souligne qu'il est encore possible d'obtenir un modèle des informations contenues dans le résidu dont le R^2 vaut 0.9589, ce qui correspond au R^2 d'un très bon modèle. Il nous indique donc qu'il est encore possible de faire un effort de modélisation.

4.3 Mesure du modèle sous-jacent

Le Gamma Test permet d'estimer une autre valeur qui, elle, est reliée au modèle "caché" derrière le bruit : la moyenne de la norme du carré du gradient $\langle |\nabla f|^2 \rangle$. En effet, le Gamma Test permet d'estimer la valeur de l'expression :

$$A(n, p) = \langle |\nabla f|^2 \rangle \cdot \langle \cos^2 \theta \rangle \quad (16)$$

Si on considère que les points $\vec{x}(i)$ sont distribués aléatoirement et uniformément, alors on peut considérer que la relation suivante est vraie :

$$A(n, p) \simeq \langle |\nabla f|^2 \rangle \quad (17)$$

Nous allons maintenant utiliser cette estimation sur le jeu de données représentant une sinusoïde bruitée. Nous allons bruitez cette sinusoïde de plus en plus et regarder si l'estimation de la moyenne de la norme du carré du gradient reste fiable.

Le jeu de données que nous construisons est le suivant :

$$y = \sin(2\pi \cdot t) + Noise \cdot rand(0, 1) \quad (18)$$

On obtient alors les résultats suivants :

Noise	$Var(r)$	Estimation de $Var(r)$	Estimation de $\langle \nabla f ^2 \rangle / 4$	R_{GT}^2
1	0.07980	0.07967	$9.59 \cdot 10^{-6}$	0.864
2	0.3493	0.3485	$7.38 \cdot 10^{-6}$	0.589
4	0.1366	1.3256	$4.28 \cdot 10^{-5}$	0.3023
8	5.2567	5.2274	$3.0 \cdot 10^{-6}$	0.089
16	21.542	21.68	$2.6 \cdot 10^{-5}$	0.021

On constate ici que le niveau de bruit a un impact sur $\langle |\nabla f|^2 \rangle$. Plus le bruit est important, plus l'impact sur $\langle |\nabla f|^2 \rangle$ est grand. Dès que la variance du bruit atteint 0.1366, la mesure de $\langle |\nabla f|^2 \rangle$ n'est plus significative car celle-ci devient négative (elle perd donc tout son sens car cette mesure doit impérativement être positive). La mesure de $\langle |\nabla f|^2 \rangle$ ne doit donc être considérée que lorsque le niveau de bruit est faible.

Nous allons maintenant utiliser cette mesure dans une application : le dimensionnement de la structure d'un réseau de neurones.

4.4 Estimation de la structure minimale d'une fonction paramétrique

Nous allons maintenant utiliser le coefficient A pour déterminer la structure minimale d'un réseau de neurones.

Dans [Koncar 97], l'auteur utilise la relation :

$$A = \frac{1}{2} \cdot \langle |\nabla f|^2 \rangle \cdot \langle \cos^2 \theta \rangle \quad (19)$$

Si θ est uniformément réparti sur l'intervalle $[0, 2\pi]$ alors $\langle \cos^2 \theta \rangle = 1/2$.

On va maintenant faire le parallèle entre la valeur de A trouvée par le Gamma Test et la valeur de A de la fonction suivante :

$$f(x_1, \dots, x_n) = \frac{1}{n} \cdot \sum_{i=1}^n \sin^2(a \cdot \pi \cdot x_i) \quad (20)$$

avec $0 \leq x_1, \dots, x_n \leq 1$. Cette fonction est représentée sur la figure 5 pour $a = 3$ et $n = 2$.

Pour cette fonction, on est capable de calculer l'expression analytique de $\langle |\nabla f|^2 \rangle$:

$$\langle |\nabla f|^2 \rangle = \frac{\int_0^1 \dots \int_0^1 \left[\left(\frac{\partial f}{\partial x_1} \right)^2 + \dots + \left(\frac{\partial f}{\partial x_n} \right)^2 \right] \cdot dx_1 \dots dx_n}{\int_0^1 \dots \int_0^1 dx_1 \dots dx_n} \quad (21)$$

Pour la fonction précédente, on trouve le résultat suivant :

$$\langle |\nabla f|^2 \rangle = \frac{a^2 \cdot \pi^2}{2 \cdot n} \cdot \left(1 - \frac{\sin(4 \cdot a \cdot \pi)}{4 \cdot a \cdot \pi} \right) \quad (22)$$

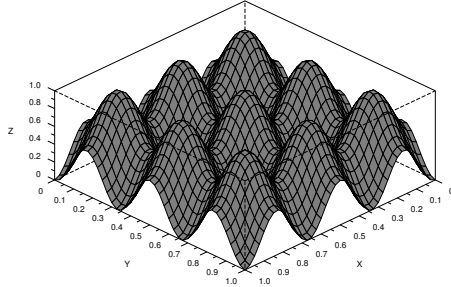


FIG. 5 – La fonction utilisée pour le dimensionnement du réseau de neurones.

Si on part de l'hypothèse que $A \simeq 1/4 \cdot \langle |\nabla f|^2 \rangle$, alors le nombre estimé de maxima locaux de la fonction vaut :

$$h = a^n = \left[\left(\frac{2 \cdot \sqrt{2}}{\pi} \right)^n n^{\frac{n}{2}} \cdot A^{\frac{n}{2}} \right] \quad (23)$$

On trouve dans [Lapedes et al. 88] une expression mettant en relation le nombre de neurones présent sur la couche cachée d'un réseau de neurones de type perceptron multicouches et le nombre de maxima locaux d'une fonction. Cette relation est la suivante :

$$n_{neurons} = (2 \cdot n + 1) \cdot h + 1 \quad (24)$$

Appliquons maintenant cette relation à l'exemple simple de la sinusoïde non bruitée. Pour ce signal, nous avons estimé une valeur de A de 10.53. On trouve alors $h \simeq 3$ et $n_{neurons} = 10$.

Nous utilisons maintenant le logiciel SNNS pour générer un réseau de neurones composé d'une entrée, de 10 neurones sur la couche cachée et d'un neurone de sortie. Pour ce réseau de neurones, nous obtenons les résultats suivants :

Réseau à 10 neurones			
n° d'essai	R^2	SSE	MSE
1	0.9852	1.85	0.0019
2	0.9861	1.74	0.0018
3	0.9960	0.5021	0.0005
4	0.9991	0.1090	0.0001

Avec :

MSE Mean of Squared Error - c'est la moyenne du carré du résidu (résidu \equiv différence entre mesure et estimation).

SSE Sum of Squared Error - c'est la somme cumulée du carré du résidu.

De plus, nous avons testé des structures de réseau de neurones à 8, 6 et 4 neurones sur la couche cachée. Nos expérimentations montrent que ces réseaux sont aussi bons que le réseau à 10 neurones sur la couche cachée (pour les réseaux à 8 et 6 neurones sur la couche cachée) et que le réseau à 4 neurones sur la couche cachée correspond à une limite. Pour ce dernier réseau, les performances sont légèrement moins bonnes. Cependant, on obtient les meilleures performances en validation sur le réseau à 10 neurones.

On peut conclure, pour cet exemple, que le Gamma Test a donné une estimation du nombre de neurones sur la couche cachée tout à fait raisonnable.

Nous allons maintenant tester l'estimation de la dimension d'un réseau de neurones sur un exemple un peu plus complexe : la fonction décrite à l'équation 20. Nous fixons le paramètre a à 3 et le paramètre n à 2. Le fichier d'apprentissage sera composé de 10201 points (x_1 et x_2 varient entre 0 et 1 par pas de 0.01). La taille du voisinage pour le calcul de l'estimation par Gamma Test est fixée à 80. On obtient la valeur suivante :

$$A = 5.4$$

Ce qui nous donne une estimation du nombre de maxima locaux de $h \simeq 8.75$ et le nombre de neurones nécessaires sur la couche cachée vaut $n_{neurons} = 45$. Les durées d'apprentissage étant beaucoup plus longues qu'avec l'exemple précédent, nous allons nous contenter de trois essais : un avec 45 neurones sur la couche cachée, un autre avec 40 neurones sur la couche cachée et un dernier avec 30 neurones sur la couche cachée.

Les résultats obtenus avec le réseau de neurones à 45 neurones sur la couche cachée sont les suivants :

Réseau à 45 neurones			
n° d'essai	R^2	SSE	MSE
1	0.9893	6.9042	0.0007
2	0.9893	6.8869	0.0007

Les résultats obtenus avec le réseau de neurones à 40 neurones sur la couche cachée sont les suivants :

Réseau à 40 neurones			
n° d'essai	R^2	SSE	MSE
1	0.9889	0.0007	7.1492
2	0.9893	0.0007	6.8990

Les résultats obtenus avec le réseau de neurones à 30 neurones sur la couche cachée sont les suivants :

Réseau à 30 neurones			
n° d'essai	R^2	SSE	MSE
1	0.9891	0.0007	7.0005
2	0.9894	0.0007	6.8426

On peut conclure, pour cet exemple, que le Gamma Test a donné une estimation du nombre de neurones sur la couche cachée tout à fait raisonnable dans le sens où le Gamma Test a estimé une structure de réseau de neurones qui a de bonnes performances en validation. On peut aussi souligner que le Gamma Test a tendance à sur-estimer la structure du réseau de neurones.

4.5 Analyse de sensibilité

Une autre application possible du Gamma Test consiste à calculer les sensibilités des entrées. Pour cela, il suffit de suivre la procédure suivante :

- calculer le R_{GT}^2 du modèle complet ;
- calculer les R_{GT}^2 des modèles avec une entrée en moins (à chaque fois, on repart du modèle complet puis on retire une entrée) ;
- on calcule ensuite la diminution relative observée sur le R_{GT}^2 calculé par le Gamma Test quand on a retiré l'entrée concernée.

Il faut bien faire attention à appliquer l'analyse de sensibilité à un modèle qui explique correctement le phénomène. En effet, si on applique l'analyse de sensibilité à un mauvais modèle (un modèle avec trop d'entrées par exemple), il est possible que la suppression d'une entrée ait pour conséquence l'augmentation du R_{GT}^2 .

Nous allons maintenant appliquer cette analyse aux jeux de données bruitées et non bruitées qui ont servi lors du test de la méthode de sélection des entrées.

- Pour le jeu de données bruitées, on obtient :

N° entrée supprimée	R_{GT}^2	Variation relative
0	-0.000593	×
1	0.000866	+246%
2	0.00225	+480%
3	0.998	+168521%

On est ici dans le cas d'un modèle qui n'est pas adapté. On a une entrée en trop et, cette entrée n'étant pas du tout significative, elle perturbe complètement le modèle. Le Gamma Test indique donc, par l'intermédiaire du R_{GT}^2 , que le modèle est très mauvais (le R_{GT}^2 est très proche de 0).

Donc, à partir du moment où cette entrée qui ne sert à rien est supprimée (cette entrée correspond à l'entrée 3), on assiste à une forte augmentation de la valeur du R_{GT}^2 . Celui-ci passe à 0.998, ce qui signifie que le modèle est très bon et nous indique par la même occasion que cette entrée ne servait très probablement à rien.

- Pour le jeu de données non bruitée, on obtient :

N° entrée supprimée	R_{GT}^2	Variation relative
0	0.998	×
1	0.533	-46.5%
2	0.502	-49.7%

Avec ce jeu de données, on est dans le cas d'application d'une analyse de sensibilité : le modèle complet est parfait, il explique bien le phénomène et le R_{GT}^2 est proche de 1. À chaque fois que l'on supprime une entrée, on voit le R_{GT}^2 fortement diminuer. Cette diminution

illustre l'importance de l'entrée dans l'explication du phénomène. Plus cette diminution sera importante, plus l'entrée sera importante dans l'explication du phénomène.
 Dans cet exemple, on constate que les deux entrées ont plus ou moins la même importance.

5 Conclusion

Nous avons vu, dans ce document, que le Gamma Test est un outil simple permettant d'estimer "rapidement" et ce, sans le calcul d'un modèle, la pertinence d'un jeu d'entrées pour expliquer un phénomène. Cet outil permet d'estimer la qualité maximum que l'on peut espérer atteindre avec une telle structure de modèle.

A partir de là, plusieurs outils peuvent être déduit de cette méthode :

- des outils de sélection d'entrées pertinentes ;
- un outils d'analyse de la sensibilité des entrées ;
- etc.

Il est possible aussi d'utiliser le Gamma Test pour arrêter l'apprentissage d'un modèle (par exemple, un modèle de réseau de neurones). En effet, ayant calculer le R_{GT}^2 du modèle correspondant au jeu d'entrées sélectionnées, on débute l'apprentissage d'un réseau de neurones et on arrête l'apprentissage quand le R_{GT}^2 du modèle atteint le R_{GT}^2 prédit par le Gamma Test sur le jeu de données d'apprentissage et avec le jeu d'entrées sélectionnées. En effet, si on cherche à poursuivre davantage l'apprentissage, le modèle va apprendre une partie du bruit et ceci aura pour conséquence une diminution de la variance du bruit de modélisation $Var(r)$. Or, on sait qu'un modèle qui apprend "par coeur" une certaine quantité de bruit aura un mauvais comportement sur un jeu de données de validation. Le modèle ne sera pas capable de "généraliser" (fournir des estimations propres sur des données qui n'ont pas servies à l'apprentissage).

La limite de R_{GT}^2 fournie par le Gamma Test semble donc correspondre à un bon critère d'arrêt de l'apprentissage d'un modèle.

Nous avons aussi vu que le Gamma Test pouvait permettre de définir la structure à proprement parler d'un modèle. Nous avons traiter l'exemple d'un réseau de neurones et vu que la structure suggérée par le Gamma Test pouvait correspondre à un bon modèle. La relation permettant de dimensionner le réseau de neurones peut certainement être transposable à d'autres modèle comme, par exemple, le modèle Lolimot.

Un bon point de départ concernant le Gamma Test est le site internet d'Antonia Jones :

<http://users.cs.cf.ac.uk/Antonia.J.Jones/GammaArchive/IndexPage.htm>

Références

- [Lapedes et al. 88] A. Lapedes, R. Farber, "How Neural Nets Work", Neural Information Processing Systems, Ed. D. Anderson, American Institute of Physics, New York, 1988.
- [Koncar 97] N. Koncar, "Optimisation methodologies for direct inverse neurocontrol", Ph. D. Thesis, University of London, Department of Computing, Imperial College of Science, Technology and Medicine, London, SW72BZ, 1997
- [Tsui 99] A.P.M. Tsui, "Smooth Data Modelling and Stimulus-Response via Stabilisation of Neural Chaos", Ph. D. Thesis, Department of Computing, Imperial College of Science, Technology and Medicine, University of London, 1999
- [Durrant 01] P.J. Durrant, "winGamma? : A non-linear data analysis and modelling tool with applications to flood prediction", Ph. D. Thesis, Department of Computer Science, University of Wales, Cardiff, Wales U.K., 2001
- [Evans 02] D. Evans, "The Gamma Test : Data derived estimates of noise for unknown smooth models using near neighbour asymptotics", Ph. D. Thesis, Department of Computer Science, Cardiff University, 2002