

ÉVALUATION DU RISQUE D'ECHEC DES ETUDIANTS DE PREMIERE ANNEE (I.N.P.S) A L'AIDE DES RESEAUX DE NEURONES. PROMOTION (99/2000)

SADI Khadidja¹

Résumé :

L'échec académique parmi les étudiants de l'université a longuement suscité de grands débats. Beaucoup de psychologues pédagogiques ont essayé de le comprendre et alors de l'expliquer. Des statisticiens ont essayé de le prévoir.

Vu l'importance de la réussite et de l'échec dans le domaine de l'enseignement supérieur qui se manifestent surtout au niveau du tronc commun et vu la nécessité de l'orientation en spécialité, le problème en question est de savoir quelles sont, parmi l'ensemble des caractéristiques présentées, celles qui expliquent la réussite de l'étudiant en première année.

Notre recherche à l'intention d'être capable de classer les étudiants en deux classes:

- la classe C1 "admis": les étudiants qui ont réussi.
- la classe C0 "ajournés": les étudiants qui ont échoué.

L'article a pour objectif l'utilisation des réseaux de neurones pour classer les étudiants en première année et de dégager les variables susceptibles d'expliquer le phénomène étudié.

INTRODUCTION

Depuis plus de deux décennies, l'enseignement supérieur traverse une crise profonde et complexe. Cette crise a conduit à une détérioration de la qualité, de la rentabilité et de l'efficacité de l'enseignement supérieur.

"Les tendances actuelles du développement économique et social indiquent que les premières années du 21ème siècle seront marquées

¹ Enseignant chercheur, chargée de cours à l'INPS

par la persistance des phénomènes tels que la mondialisation de l'économie et du commerce, la démocratisation des régimes politiques et l'évolution rapide des connaissances et des technologies. Ces phénomènes qui auront des répercussions directes sur le marché de l'emploi, exigent des réformes conceptuelles et institutionnelles adéquates de l'organisation et de la gestion de l'enseignement supérieur"¹.

Pour cela on s'interroge sur la qualité et l'efficacité du système éducatif, ce qui renvoie à des analyses en termes d'évaluation. Cela vise à rechercher les bases du bon fonctionnement du système éducatif.

L'échec académique parmi les étudiants de l'université a longuement suscité de grands débats. Beaucoup de psychologues pédagogiques ont essayé de le comprendre et alors de l'expliquer. Des statisticiens ont essayé de le prévoir.

Vu l'importance de la réussite et de l'échec dans le domaine de l'enseignement supérieur qui se manifestent surtout au niveau du tronc commun et vu la nécessité de l'orientation en spécialité, le problème en question est de savoir quelles sont, parmi l'ensemble des caractéristiques présentées, celles qui expliquent la réussite de l'étudiant en première année.

Notre recherche a l'intention de classer les étudiants dans deux classes:

- la classe C1 "admis" : les étudiants qui ont réussi.
- la classe C0 "ajournés": les étudiants qui ont échoué.

L'article a pour objectif l'utilisation des réseaux de neurones pour classer les étudiants en première année et de dégager les variables susceptibles d'expliquer le phénomène étudié.

L'article est divisé en trois parties. La première partie présente les données utilisées, la deuxième présente le modèle neuronal conçu et la troisième partie présente les résultats obtenus.

¹ FORUM DES ASSOCIATIONS DES ETUDIANTS EN AFRIQUE SUR L'ENSEIGNEMENT SUPERIEUR AU 21ème SIECLE Accra (Ghana) 23-25 mars 1998.

Il faut noter, par ailleurs que cette étude reste limitée du fait que nous ne disposons pas d'informations sur l'environnement social de l'étudiant.

Logiciels utilisés :

- Excel 2003.
- Le logiciel mathématique MATLAB version 7.

1. LES DONNEES

Notre étude concerne les étudiants de l'INPS¹ en cycle d'ingénieurs. La promotion étudiée est constituée d'une population mère de 461 étudiants inscrits pour la première fois durant l'année universitaire 1999/2000.

Les échantillons.

L'échantillon de base retenu, pour l'étude, est de taille 100 (étudiants). Quand on procède à la validation croisée du modèle, cet échantillon est réparti en trois échantillons :

- L'échantillon d'apprentissage de taille 40 (20 pour la classe C1 et 20 pour la classe C2). Il sert à construire les modèles.
- L'échantillon de validation de taille 40 (20 pour la classe C1 et 20 pour la classe C2). Il sert à définir l'architecture du réseau.
- L'échantillon de test de taille 20 (10 pour la classe C1 et 10 pour la classe C2). Il sert à tester la capacité de généralisation du réseau.

Les variables.

L'évaluation de l'échec de l'étudiant de première année repose ici sur les informations recueillies des fichiers archives.

Pour chaque étudiant on observe à l'entrée des caractéristiques pédagogiques et des caractéristiques non pédagogiques.

Ainsi, les variables retenues sont de deux types :

- Les variables quantitatives :
 1. la note de mathématique obtenue au baccalauréat (*math*).
 2. la moyenne obtenue au baccalauréat (*moyb*).
 3. la note de français obtenue au baccalauréat (*fran*).
 4. l'âge de l'étudiant (*âge*).
- Les variables qualitatives :
 1. le sexe de l'étudiant (*sexe*).

¹ I.N.P.S : Institut National de la Planification et de Statistique.

2. la série du baccalauréat (sciences de la nature, sciences exactes, science de gestion, techniques comptables) (*série*).
3. la région d'origine (Alger, hors d'Alger) (*ro*).

2. LES RESEAUX DE NEURONES

Conçus à l'origine comme des modèles mathématiques très simplifiés de notre cerveau, les réseaux de neurones formels ou artificiels (*R.N.A*) sont des systèmes d'apprentissage par exemples, imitant en cela la façon dont nous apprenons par essais, erreurs et corrections successives à effectuer des tâches complexes.

Concrètement, ces techniques neuronales sont des outils mathématiques et algorithmiques qui s'avèrent être puissants et commodes pour résoudre des problèmes non linéaires et complexes.

Le domaine s'est développé extrêmement vite en proposant des méthodes pratiques pour résoudre des problèmes de grande taille et des applications ont été réalisées dans la plupart des domaines traditionnels de l'ingénierie (diagnostic, contrôle, reconnaissance de caractères et d'images, prévision de série temporelles, discrimination, etc.).

Les réseaux de neurones artificiels sont donc des modèles mathématiques, se présentant comme un ensemble d'unités de calcul fortement connectés appelé neurone formel ou artificiel. A ce titre ils peuvent être décrits par leur architecture et leurs composants.

2.1 Présentation du réseau de neurones :

Nous avons utilisé un réseau multicouche (figure 1) composé des éléments suivants :

- Une couche d'entrée qui reçoit des variables explicatives x_i .
- Une couche de sortie qui contient un seul neurone (neurone de décision) avec une valeur binaire représentant la classe à discriminer.
- Une couche cachée constituée de N^1 neurones. Ce nombre est atteint progressivement après plusieurs tests sur l'échantillon d'apprentissage jusqu'à atteindre l'optimum d'apprentissage et de généralisation.

¹ N est fixé lors de la validation du modèle (voir tableau validation).

Deux fonctions de transfert sont utilisées :

- La fonction logistique pour la couche de sortie.
- La fonction tangente hyperbolique pour la couche cachée.

L'apprentissage d'un tel réseau est supervisé et utilise l'algorithme de rétro-propagation du gradient de l'erreur.

2.2 Mise en œuvre du réseau :

Pour optimiser la mise en œuvre du réseau on contrôle le choix de plusieurs éléments :

1. les paramètres d'apprentissage. On fixe le seuil maximal d'itération, le taux d'apprentissage et la valeur du moment.
2. le choix d'un jeu de variables initiales. Les variables forment les entrées du réseau, on propose plusieurs scénarios ou modèles (chaque modèle est un programme) selon le jeu de variables choisi.
3. le nombre de neurones sur la couche cachée. Pour chaque modèle on présente les mêmes exemples et on modifie progressivement le nombre de neurones sur la couche cachée jusqu'à convergence du réseau vers l'erreur désirée.

Pour tous ces choix on utilise la validation croisée¹.

On dispose de sept (7) variables. On se propose plusieurs choix initiaux (*tableau 1*). Le nombre optimal de neurones dans la couche cachée est déterminée selon les critères de choix du modèle (maximisation du pourcentage de bon classement et minimisation de l'erreur).

¹ *Validation croisée* : le principe est de définir l'arrêt de l'apprentissage tout en évitant le sur apprentissage. Le principe est de scinder l'échantillon de base en deux: échantillon d'apprentissage et échantillon de test. Une fois déterminé un réseau intéressant, on l'applique sur des échantillons tests indépendants de l'échantillon de base.

Tableau 1 : Tableau de validation

Nom du réseau	Nombre de variables	N	Epochs	Taux de bon Classement	Variables utilisées
Réseau 1	3	7	6340	70%	<i>math,moyb,fran</i>
Réseau 2	4	10	7164	55%	<i>math,moyb,fran,age</i>
Réseau 3	5	13	3556	63%	<i>math,moyb,fran,age,serie</i>
		15	4044	68%	<i>math,moyb,fran,age,sexe</i>
		17	6587	63%	<i>math,moyb,fran,age,or</i>
Réseau 4	6	10	4737	75%	<i>math,moyb,fran,age,serie,or</i>
		14	1476	75%	<i>math,moyb,fran,age,serie,sexe</i>
		20	4158	70%	<i>math,moyb,fran,age,sexe,or</i>
Réseau 5	7	16	1596	77,50%	<i>Toutes les variables</i>

L'échantillon d'apprentissage ou de construction du modèle neuronal est présenté plusieurs fois au réseau de manière itérative jusqu'à convergence de l'erreur vers un minimum fixé au préalable (fig2).

Le réseau 1 donne un bon taux de classement à partir des trois variables (tableau 1). Ce qui implique qu'elles expliquent bien le phénomène étudié. On rajoute progressivement les autres variables : une, puis deux et enfin trois en étudiant toutes les combinaisons possibles entre elles.

D'après ce tableau, le meilleur résultat est donné par le réseau 5(toutes les variables sont indispensables pour l'étude du phénomène). Néanmoins, pour la suite nous retiendrons les trois modèles : le réseau 1, le réseau 4, et le réseau 5 car ils donnent des résultats intéressants.

3. LES RESULTATS

Dans cette phase on présente au réseau des exemples non connus et on teste sa capacité de généralisation et donc sa robustesse de prévision.

TABLEAU DES RESULTATS

Modèle	N (nb neurones de la couche cachée)	Clas se	Taux de bon	Taux de mauvais
Réseau 1	7	Clas se C0	60%	40%
		Clas se C1	40%	60%
		total	50%	50%
Réseau 4	14	Clas se C0	50%	50%
		Clas se C1	60%	40%
		total	55%	45%
Réseau 5	16	Clas se C0	60%	40%
		Clas se C1	60%	40%
		total	60%	40%

CONCLUSION

Au terme de cet article il apparaît que les réseaux de neurones artificiels donnent aujourd'hui de bons résultats dans les applications. Il semblerait que ce soit dans le domaine de la classification, en particulier.

Cependant, ils présentent encore certaines limites : la construction du réseau, le retraitement des variables d'entrée et le réglage des paramètres d'apprentissage relèvent encore trop de l'intervention humaine. Par ailleurs ils n'ont encore qu'un pouvoir explicatif réduit. Mais tous ces problèmes font aujourd'hui l'objet de recherches dont

on peut espérer qu'elles apportent des réponses satisfaisantes dans les années qui viennent. Mais, comme le signale Hoptroff (1993), si l'outil est intéressant, comme toutes les méthodes de prévisions, les réseaux de neurones ne pourront jamais prévoir l'imprévisible.

BIBLIOGRAPHIE

- 1- Analyse de la performance des étudiants ingénieurs de l'INPS de la promotion 1999/2000 via les modèles logit binaire et logit multinomial séquentiel.
- 2- Michel Vernieres ed.économica 1995. Ajustement éducation emploi.
- 3- Heckman, James J., Lance Lochner et Christopher Taber (1998a), « Tax Policy and Human- Capital Formation », *American Economic Review*, vol. 88, no. 2 (mai), pp. 293-297.
- 4- Heckman, James J., Lance Lochner et Christopher Taber (1998b), « Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents », *Review of Economic Dynamics*, vol. 1, no. 1 (Janvier), pp. 1-58.
- 5- Heckman, James J., Lance Lochner et Christopher Taber (1999), « Human Capital Formation and General Equilibrium Treatment Effects: A Study of Tax and Tuition Policy », *Fiscal Studies*, vol. 20, no. 1 (mars), pp. 25-40.
- 6- Heckman, James J. (2001), « Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture », *Journal of Political Economy*, vol. 109, no. 4, pp. 673-748.
- 7- Lemieux, Thomas et David Card (2001), « Education, Earnings, and the 'Canadian G.I. Bill' », *Revue canadienne d'économique*, vol. 34, no. 2 (mai), pp. 313-344.
- 8- Mincer, Jacob (1974), *Schooling, Experience, and Earnings*, New York, Columbia University Press. Sauer, Robert M. (1998), «

Job Mobility and the Market for Lawyers », *Journal of Political Economy*, vol. 106, pp. 147-171.