

Apports de la segmentation de la population des dirigeants au processus de gestion des ressources humaines : cas de l'administration algérienne

By/Par | Rédha Tir

Université Alger II Redha_tir@yahoo.com

ABSTRACT

From the results of a survey conducted in 2005 on a sample of 203 public officials, a typology of three fuzzy profiles has been cleared to design an allocation system based on their leadership styles, work ethics and human resource management concerns. Based on this typology fuzzy techniques borrowed from artificial intelligence have been used to learn classification rules. These techniques are four in number: the neural network, the genetic algorithm, the decision tree, and the theory of rough sets. The results of the study and its perspectives will be presented and discussed throughout this paper.

Keywords: Potential detection, leadership, segmentation, manager profile, artificial intelligence.

RÉSUMÉ

A partir des résultats d'une enquête effectuée en 2005 sur un échantillon de 203 dirigeants publics, une typologie floue de trois profils a été dégagée en vue de concevoir un système d'affectation des dirigeants en fonction de leur style du leadership, sens du travail, et leurs préoccupations de gestion des ressources humaines. En se basant sur cette typologie floue, des techniques empruntées à l'intelligence artificielle ont été appliquées pour apprendre des règles de classification. Ces techniques sont au nombre de quatre : le réseau neuronal (Neural Network), l'algorithme génétique (Genetic Algorithm), l'arbre de décision (Decision Tree) et la théorie des ensembles approximatifs (Rough Sets). Les résultats de l'étude ainsi que ses perspectives seront présentées et discutés tout au long de cette communication.

Mots clés : détection de potentiel, leadership, segmentation, profil du dirigeant, intelligence artificielle.

Classification JEL: C45, D73

INTRODUCTION

Les organisations modernes recourent de plus en plus aux systèmes d'aide à la décision en vue d'extraire le potentiel de leurs membres. En gestion des ressources humaines (GRH), l'engagement à l'égard de l'organisation, la performance au travail et le développement personnel sont des facteurs déterminants du succès ou de l'échec du système de gestion en place. Donc le recrutement des compétences doit se faire conformément à un certain nombre de critères objectifs, équitables et transparents¹.

La sélection des dirigeants constitue un enjeu important dans les organisations contemporaines. En effet, l'administration publique, garant de l'intérêt général, est appelée à se moderniser sur le plan du ciblage et de la sélection de ses dirigeants. En fait, ces dirigeants doivent être dynamiques, enthousiastes et énergiques². Ces qualités permettent de motiver les subalternes, communiquer leurs idées, et de les intéresser à atteindre des objectifs stratégiques à long terme. En pratique, la sélection des dirigeants se fait, dans la plupart des cas, sur des bases subjectives qui ne tiennent pas en compte des critères d'efficacité, d'efficience ou encore de qualité. Les conséquences négatives sont très importantes du moment où l'administration continue à fonctionner selon un modèle fait d'autorité, de lourdeur et de secret (Taïb, 1996). L'objectif de cette étude est d'établir un modèle permettant d'aider l'autorité de nomination à sélectionner les meilleurs candidats à une fonction ou un poste supérieurs, et à identifier leurs besoins en formation.

A partir des résultats d'une enquête menée en 2005, une typologie floue a été établie sur la base d'un ensemble de mesures psychologiques. Ensuite, quatre techniques relevant du Soft Computing ont été utilisées pour répondre au besoin de cette étude. Le Soft Computing regroupe l'ensemble de techniques empruntées à la biologie et la physique appliquée en vue de simuler le raisonnement humain, et par voie de conséquence, faire face aux limites des modèles statistico-mathématiques quant à l'imprécision et l'ambiguïté des données. Pour effectuer la segmentation, six attributs relatifs au dirigeant ont été choisis ainsi que le nombre de segments retenus est de trois.

Les différentes techniques utilisées visent la découverte de relations non linéaires entre ces variables.

Enfin, elles permettent de générer un ensemble de règles de classification aisément interprétables. En amont, cette typologie permettra de connaître les points forts et les faiblesses des candidats aux fonctions et postes supérieurs à travers leurs attitudes envers le travail administratif, les relations interpersonnelles dans l'administration et enfin envers leurs préoccupations de gestion des ressources humaines (GRH). En aval, elle contribue à l'identification des besoins en formation et à la personnalisation des programmes. Ceci dit, la typologie élaborée est dépendante de la culture administrative qui caractérise les services publics.

¹ - Taïb E. 1996. La Réforme de l'administration en Algérie. *Annuaire de l'Afrique du Nord*. Tome XXXV. CNRS éditions, pp. 343-379.

Liu W., Lepak D. P., Takeuchi R. and Sims H. R. 2003. Matching leadership styles with employment modes: strategic human resource management perspective. *Human Resource Management Review*, 13: 129-130.

DIMENSIONS DU PROFILE DU DIRIGEANT PUBLIC

La gestion stratégique des ressources humaines (GSRH) dans le service public suppose l'existence d'un système de détection et de sélection des hauts potentiels. Le profil des dirigeants constitue, à cet égard, un cadre adéquat permettant de cerner les principaux aspects de leur comportement quotidien dans l'administration. En réalité, ces aspects reflètent l'ensemble des compétences qu'on peut exiger à un candidat à une fonction supérieure ou un poste supérieur.

Pour déterminer les facteurs reliés au comportement du dirigeant dans la fonction publique algérienne, des entretiens directifs et semi-directifs, à propos de la GSRH, ont été conduits afin de générer un ensemble d'items de questionnaire concernant les différents aspects et caractéristiques du profil du dirigeant.

Ainsi, la littérature sur ce sujet démontre qu'en théorie le profil du dirigeant se décline en plusieurs compétences telles que le style du leadership, les comportements dits de citoyenneté organisationnelle, et bien entendu, son rôle dans l'administration.

Le cadre théorique et conceptuel de cette étude suppose l'existence de trois facettes du comportement du dirigeant qui ont un effet médiateur et d'ajustement entre sa perception du sens du travail et ses préoccupations de gestion des ressources humaines (GRH). Le sens du travail couvre les aspects de l'utilité, l'éthique, l'efficacité, les conditions de réussite d'un dirigeant, le plaisir au travail, et enfin la sécurité personnelle. Ainsi, Le comportement du dirigeant se compose de plusieurs dimensions qui sont indiquées dans la figure 1. En dernier ressort, les préoccupations de GRH comportent deux niveaux. Le premier représente les différentes politiques de GRH au sein de l'administration telles que : le développement du personnel, la planification, les systèmes de rémunération, la sélection et la communication. Le deuxième niveau comporte les diverses pratiques de GRH exercées quotidiennement par un dirigeant public.

Sur la base des entretiens et du modèle théorique, un questionnaire a été élaboré et administré à un échantillon de dirigeants publics en poste. Ce questionnaire nous donne des informations sur les aspects du sens du travail, les caractéristiques du leadership existant, les responsabilités et les comportements discrétionnaires, ainsi que les diverses manières de gérer le capital humain disponible.

Afin de mieux cerner ces variables, des échelles de mesure ont été construites sur la base de la traduction renversée d'un ensemble d'échelles connues dans la littérature. De plus, un effort personnel a été dispensé en vue de générer des items supplémentaires et de finaliser le questionnaire. Par ailleurs, des analyses factorielles (exploratoire et confirmatoire) ont été effectuées dans le but de valider les échelles de mesure sus-indiquées, et de tester les hypothèses du modèle. Enfin, son application dépend dans une large mesure de la volonté de l'autorité de nomination.

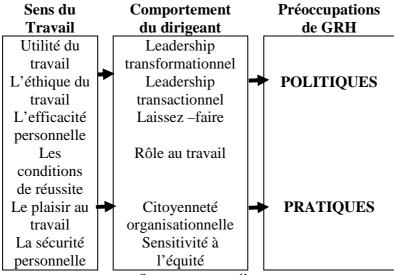


Figure 1. Modèle de recherche

Source : conçu par l'auteur

Échantillon

L'échantillon d'apprentissage est composé de 203 dirigeants publics et couvre plusieurs postes et fonctions supérieurs institués en Algérie (Ex: secrétaire général, directeur général, etc.).

Au total, 23 variables, ont été choisies en vue de mieux classer les dirigeants dans des segments homogènes.

La collecte de données à été effectuée sur trois sites différents où le questionnaire a été administré auprès de 150 dirigeants inscrits dans des programmes de formation continue. Le reste (53) a été contacté directement par l'auteur. L'enquête s'est déroulée entre le 10 septembre 2004 et le 11 juin 2005.

La population constituant l'échantillon est de sexe masculin à (86.7%) et féminin à (13.3%). L'âge moyen des participants à l'enquête est de 44 ans. Leur ancienneté moyenne dans la fonction ou le poste est de 4,5 ans. Les fonctions représentées sont : les secrétaires généraux (5.9%), Les directeurs généraux (1%), les directeurs (21.2%), les sous-directeurs (28.1%), et des chefs de service (43.8%). Ces dirigeants travaillent dans des administrations centrales (16.7%), déconcentrées (38.9%), locales (10.8%) et dans des établissements publics (33.5%). Enfin, le niveau d'instruction varie du postuniversitaire (11.3%) au secondaire (21.7%) passant par le niveau universitaire (64%) et d'autres diplômes équivalents (3%).

Algorithme de segmentation.

Dans le but d'établir une typologie de profils des dirigeants, le processus de segmentation comporte deux étapes.

La première consiste en la découverte du nombre de segments à définir. L'objectif serait de mieux connaître les besoins et exigences de chaque fonction ou poste. L'opposition éventuelle entre la taille de chaque segment et le degré de son homogénéité est prise en compte.

A cet égard, pour connaître le nombre probable de segments composant l'échantillon, l'algorithme des cartes auto-organisatrices SOM (Self Organizing Maps) dit aussi de Kohonen a été utilisé en vue de représenter les segments recherchés. Ces cartes représentent des réseaux de neurones (voir plus loin) à apprentissage non supervisé. Elles sont composées d'une grille de neurones (nœuds, unités) auxquels seront présentés des stimuli. Un stimulus est un vecteur de dimension définie d qui décrit un objet à classer. Les cartes permettent le classement d'objets (individus) sur lequel aucune information relative à leur classification n'est disponible à priori. Une fois le réseau organisé, chaque objet est classé dans le segment qui lui correspond. Ainsi, l'utilisation des cartes de Kohonen dans cette étude consiste à présenter les 23 variables (inputs) caractérisant la perception du sens du travail, le comportement au travail, et les préoccupations de GRH chez le dirigeant public à la carte pour représenter les individus en segments.

Ensuite, elle représente le classement sous forme graphique compréhensible à l'être humain et qui permettra par la suite de découvrir le nombre de segments représentant la population étudiée.

L'exécution de l'algorithme SOM sous Excel 2003 démontre que le nombre de segments est de trois (3). Par voie de conséquence, ce nombre a été retenu pour la suite de l'étude.

La deuxième étape vise l'obtention des degrés d'appartenance de chaque observation (dirigeant) à un segment ou plus. Pour ce faire, nous avons utilisé l'algorithme FCM (Fuzzy C-Means Clustering) fonctionnant sous MATLAB 7 (The Mathworks, 2004).

Cet algorithme produit des degrés d'appartenance à un nombre de segments fixé préalablement comme paramètre. Les principes de la logique floue représentent la base du fonctionnement de cet algorithme. Ses résultats constituent un point de démarrage pour l'élaboration d'une typologie dite « floue » Dans le cas de cette étude, le degré d'appartenance le plus élevé parmi les trois résultants a été retenu pour caractériser l'affectation de chaque dirigeant au segment qui lui correspond. Le but recherché demeure la constitution d'un modèle capable de générer des règles de classification à partir des résultats issus de la typologie floue. Ces règles seront, par la suite, utilisées comme outil de prédiction des classes auxquelles peut appartenir chaque nouveau candidat à une fonction ou un poste supérieur, et ce dans le but d'aider l'autorité de nomination à reconnaître le profil qui répond le mieux aux compétences dudit candidat.

.

³ Nguyen Phuong T., Cliquet G. et Leroy F. 2003. L'opposition entre la taille du marché et degré d'homogénéité des segments : approche par la logique floue. *Décision Marketing*, N° 32 : 56.

Typologie résultante

Après analyse des résultats quantitatifs et qualitatifs de l'enquête, trois profils, composant la typologie floue, ont été dégagés et décrits sur la base des réponses obtenues.

Tableau 1. Profils des Dirigeants

DIMENSION	Profil 1	Profil 2	Profil 3
Sens du Travail	Moyen	Faible	Moyen
Sensitivité à l'équité	Moyenne	Elevée	Moyenne
Leadership transformationnel	Élevé	Faible	Élevé
Leadership Transactionnel	Élevé	Élevé	Élevé
Laissez – Faire	Faible	Faible	Faible
Rôle au travail	Satisfaisant	Satisfaisant	Satisfaisant
Citoyenneté Organisationnelle	Faible	Faible	Moyenne
Préoccupations de GRH	Moyennes	Faibles	Faibles

Ainsi, l'échantillon comprend donc les 203 interviewés décrits par :

6 attributs socio-professionnels comportant au total 26 modalités. Ces attributs sont : le sexe, l'âge, le niveau d'instruction, l'ancienneté dans la fonction ou le poste, la fonction ou le poste, et enfin l'administration employeuse.

Le segment auquel le dirigeant appartient, qui correspond au degré d'appartenance le plus élevé des trois produits par le FCM.

Cet échantillon a été divisé en deux groupes, l'un servant à l'apprentissage et l'autre au test des règles obtenues.

Mais avant de réaliser les calculs, il est nécessaire de donner un aperçu des différentes techniques de segmentation préconisées pour cette étude.

Ces techniques sont au nombre de quatre. Dans un premier temps, les réseaux neuronaux seront présentés, suivis d'une description des algorithmes génétiques et leur fonctionnement. Ensuite, en vue d'évaluer la discrimination entre les trois segments concernés, deux autres méthodes ont été utilisées : l'arbre de décision et la théorie des ensembles approximatifs.

MÉTHODES DE SEGMENTATION

Il existe plusieurs approches pour traiter ce problème. En effet, les règles de classification apprises à partir de l'échantillon d'apprentissage doivent être :

Complètes : toutes les observations de cet échantillon doivent être couvertes par au moins une règle ;

Correctes: ces règles doivent en outre classer ces observations (individus) correctement;

Concises : le nombre de règles résultantes doit être minimisé.

De plus, ces règles doivent être aisément interprétables en vue de faciliter la tâche de l'autorité de nomination face aux nouveaux candidats aux postes de commandement de l'administration publique.

a) Réseaux de neurones : présentation et évaluation

Les réseaux neuronaux revêtent une importance cruciale dans plusieurs domaines. La psychologie et la GRH n'échappent pas à l'emploi de ce type de méthodes ou techniques modernes. Il s'agit d'un outil issu de l'intelligence artificielle, habituellement utilisé en sciences appliquées (biologie, physique, ... etc.) et qui a fait son entrée en sciences sociales. Le réseau de neurones est utilisé à des fins de prévision, de classification et de reconnaissance de forme en général.

Selon Paquet⁴, il existe deux raisons essentielles qui poussent les chercheurs à s'intéresser à cet outil. Primo, contrairement aux techniques statistiques classiques, le réseau neuronal ne nécessite aucune hypothèse sur les variables. Secondo, il représente un instrument adapté pour traiter des problèmes complexes et non structurés, d'où l'impossibilité de spécifier, à priori, la forme de la relation entre les variables étudiées.

Globalement, le réseau de neurones peut être utilisé pour différentes questions. On pourrait citer : la segmentation, la détection des entreprises en difficulté, la gestion de portefeuille, la prévision des séries financières, du taux de change, l'évaluation d'actifs et le choix de stratégies.

b) Architecture du réseau multicouche

Le réseau à couches est le plus utilisé dans ce domaine. Il est organisé, comme son nom l'indique, en couches. Chaque couche comporte plusieurs neurones. Chaque neurone représente une unité de calcul autonome reliée aux neurones de la ou les couches précédentes.

_

⁴ Paquet P. 1997. L'utilisation des réseaux de neurones artificiels en finance. Document de recherche n°1. Laboratoire Orléanais de Gestion.

⁵ Bolgot S. et Meyfredi J. C. 1999. Réseaux de neurones, lissage de la fonction d'actualisation et prévision des OAT démembrées : une étude empirique. Document GREQAM, Université de la Méditerranée, France.

Montagno R., Sexton R. S. and Smith B. N. 2002. Using neural networks for identifying organizational improvment strategies, www.faculty.smsu.edu/r/rss000f.

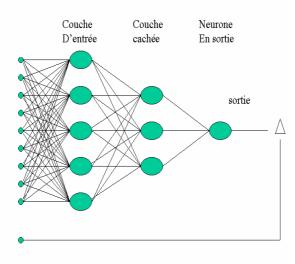


Figure 2. Architecture d'un réseau multi-couches

Source: conçu par l'auteur

Le réseau que représente la figure 2 comporte trois couches. La couche d'entrée (inputs), une couche de sortie comportant un seul neurone et donnant le résultat de tous les calculs internes.

Il existe entre ces deux couches une couche non visible de l'extérieur (hidden) appelée couche « cachée », qui est, en fait, une boite ou engin noir. Là où tous les calculs intermédiaires et transformations s'effectuent.

Quant au fonctionnement du réseau, l'algorithme d'apprentissage aura pour tâche d'évaluer des poids dits synaptiques, qui relient les neurones entre eux. Chaque neurone reçoit les informations fournies par les neurones de la couche précédente. Il calcule, ensuite, son potentiel d'activation. Une fonction d'activation sert à déterminer l'impulsion à envoyer aux neurones de la couche qui suit afin de calculer le potentiel de sortie (outputs).

c) Apprentissage du réseau de neurones

Afin que le réseau puisse découvrir la forme de la relation entre les variables, il suit, en général, deux types d'apprentissage. Le premier, dit supervisé consiste en l'existence d'un échantillon sur lequel le réseau apprend (s'entraîne) à reconnaître les formes. L'apprentissage non supervisé, qui est le second type, est utilisé lorsque on n'est pas en mesure de présenter au système un échantillon mettant en regard une quantité d'information, et la forme qu'elle est censée représenter.

Paquet P. 1997. L'utilisation des réseaux de neurones artificiels en finance. *Document de recherche* $n^{\circ}I$. Laboratoire Orléanais de Gestion.

Par voie de conséquence, le réseau s'auto organise, comme c'est le cas pour l'algorithme de Kohonen, en vue d'être capable de découvrir la forme à partir des données fournies et sans aide extérieure.

Les données disponibles doivent être réparties en trois sous-ensembles (apprentissage, validation et test) à raison de 60%, 30% et 10% respectivement.

A partir de l'échantillon d'apprentissage, le réseau de neurones se paramètre. Autrement dit, l'algorithme d'apprentissage ajuste les poids synaptiques tout en minimisant une fonction de coût. Cette dernière n'est, en fait, que la somme des carrées des erreurs produites par le réseau eu égard le résultat souhaité.

La rétro-propagation des erreurs se fait continuellement jusqu'à ce que la fonction de coût soit minimisée, ou jusqu'à ce que le concepteur intervienne pour y mettre fin.

Pour éviter le sur ajustement « overfitting », une procédure d'early stopping devrait être exécutée. Elle consiste en l'introduction de l'échantillon de validation. Le point où cet échantillon réalise le minimum d'erreur reflète la meilleure performance.

d) L'évaluation du réseau de neurones

L'échantillon d'apprentissage servira au paramétrage. Le deuxième servira à la validation (arrêter l'apprentissage quand le niveau d'erreur est le plus bas possible). Autrement dit, l'objectif de celui-ci est d'arrêter le processus d'apprentissage lorsqu'il semble donner un résultat satisfaisant en minimisation de la fonction d'erreur.

Le troisième, sera réservé pour évaluer les capacités du réseau à se généraliser et à simuler les outputs relatifs à un autre ensemble de données.

Enfin, il est à signaler que le réseau neuronal est privilégié pour traiter des non linéarités, et de la complexité en travaillant sur des données caractérisées par l'incomplétude et l'imprécision.

De même, les réseaux de neurones permettent le traitement des variables qualitatives à travers des neurones recevant des valeurs binaires comme c'est le cas pour la segmentation et la classification en sciences comportementales.

L'algorithme génétique (AG)

Holland fut le premier qui a proposé les Algorithmes Génétiques (AGs) dans les années 70. Ces derniers sont des algorithmes d'optimisation stochastique fondés sur les mécanismes de l'évolution génétique des espèces, plus précisément, du principe de sélection naturelle.

a) Présentation des algorithmes génétiques

Les AGs travaillent sur une population de points au lieu d'un point unique. Contrairement aux autres méthodes, Ils utilisent un codage des paramètres et non les paramètres eux

⁸ Collins J. M. and Clarck M. R. 1993. An application of the theory of neural computation to the prediction of workplace behaviour: an illustration and assessment of network analysis. Personnel Psychology, 46: 503-524.

mêmes. De plus, les AGs ne prennent en considération que les valeurs de la fonction étudiée, pas sa dérivée, ou une autre connaissance auxiliaire⁹.

Lors de leur fonctionnement, ils utilisent des règles de transition probabilistes (en situation de croisement ou de mutation par exemple) et non déterministes.

Un AG manipule une population de taille constante N. Cette population se compose d'individus, chacun représente le codage d'une solution potentielle au problème posé, donnée sous la forme d'une chaîne de caractères. Chaque chaîne de caractères correspond à un chromosome (individu/séquence). Chaque caractère est un gène et chaque lettre de l'alphabet est un allèle.

Le locus est la position d'un gène au sein d'un chromosome. Reste la fonction sélective, qui permet d'associer une valeur à chaque individu de la population. Cette fonction est souvent une transformation de la fonction objective à optimiser (appelée fonction de fitness).

Par ailleurs, il existe trois principaux types de codage : binaire, gray ou réel. Le fonctionnement d'un algorithme génétique se base sur les phases suivantes 10 :

- Initialisation : il s'agit de générer aléatoirement une population d'individus de taille donnée ;
- Evaluation : chaque chromosome est décodé puis évalué ;
- Sélection : utilisation d'une technique de sélection appropriée afin de créer une nouvelle population de N chromosomes ;
- Reproduction : il s'agit, en fait, de recombiner deux individus appariés (lors de la phase précédente) pour créer deux nouveaux individus. Il y a donc possibilité de mutation ou de croisement au sein de la nouvelle population.
- Retour : à la phase de décodage et d'évaluation des chromosomes, jusqu'à l'arrêt du processus.

b) Opérations génétiques classiques

Ces opérations jouent un rôle déterminant dans la réussite d'un AG. Les principaux opérateurs sont en nombre de trois :

- Opérateur de Sélection : la sélection est la première étape du fonctionnement d'un algorithme génétique. L'objectif est de sélectionner des chromosomes en fonction de leur valeur sélective. Les individus qui disposent d'une meilleure valeur sélective seront choisis. Pour ce faire, il existe plusieurs méthodes pour la sélection. La roue de loterie biaisée (Roulette Wheel) de Goldberg en est la plus connue et utilisée.
- Opérateur de Croisement : la recombinaison se fait en deux étapes essentielles :

.

Ordon O., Gomide F., Herrera F., Hoffmann and Magdalena L. 2004. Ten Years of Genetic Fuzzy Systems: Current framework and News Trends. Fuzzy Sets and Systems, 141: 5-31.

¹⁰ Cordon O., Gomide F., Herrera F., Hoffmann and Magdalena L. 2004. Ten Years of Genetic Fuzzy Systems: Current framework and News Trends. *Fuzzy Sets and Systems*, 141: 5-31.

L'appariement : consiste à choisir deux individus parmi ceux qui ont été sélectionnés pour en créer d'autres (Cordon et al, 2004). Il y a diverses méthodes pour effectuer cette opération, parmi lesquelles : l'appariement aléatoire, la consanguinité et le croisement entre lignées ou inter fécondations (la lignée résulte de l'appariement de deux individus proches au sens de la distance de Hamming).

Le croisement : cet opérateur combine deux individus appariés avec une probabilité Pc, qui est classiquement, comprise entre 0.5 et 0.9. Plus cette probabilité est élevée plus la population subira un changement.

- Opérateur de Mutation : lors d'un processus de sélection, certaines informations peuvent disparaître de la population. Afin d'éviter l'établissement de populations uniformes incapables d'évoluer, l'opérateur de mutation permettra de modifier aléatoirement, avec une probabilité donnée Pm, la valeur d'un composant de l'individu, mais il ne crée, généralement, pas de meilleurs individus. Le choix de cette probabilité est d'une importance cruciale. En effet, Pm dépend de la taille N de la population, et de la longueur des individus (chromosomes). Elle est inversement proportionnelle à la probabilité de croisement ; plus elle est élevée plus la performance de l'AG décroît.

Actuellement, les chercheurs visent une réduction du temps de convergence de l'AG, et l'amélioration de la qualité de la population finale.

L'arbre de décision.

L'arbre de décision est une méthode qui a pris essentiellement son essor dans le cadre des approches d'apprentissage automatique (machine learning) en intelligence artificielle. Il est utilisé à des fins de prédiction notamment l'affectation d'observations ou d'objets à des segments de variables dépendantes catégorielles à partir de leurs mesures sur une ou plusieurs variables indépendantes.

Cette technique peut être utilisée lorsque les techniques statistiques connues échouent comme elle peut être considérée comme méthode exploratoire. L'un de ses atouts c'est qu'elle ne dépend pas de la distribution des données collectées.

L'élaboration d'un arbre de décision passe par deux phases. La première représente la construction de l'arbre. Sur la base d'un échantillon d'apprentissage, un processus récursif de division de l'espace des données en sous espaces de plus en plus pures en terme de classes sera déclenché. L'estimation se fera sur la base d'un critère de prévision.

La deuxième phase consiste en l'élagage (pruning) de l'arbre. Il s'agit de supprimer les branches (parties terminales) peu représentatives pour atteindre de bonnes performances prédictives.

Cette opération nécessite un critère pour désigner les branches à élaguer. Après élagage, les nouvelles feuilles seront labellisées sur la base de la distribution des observations utilisées dans le processus d'apprentissage.

Afin d'évaluer la performance de l'arbre, on peut recourir à plusieurs méthodes :

- L'utilisation d'un échantillon test ;
- La validation croisée;

- L'estimation statistique.

Ainsi, l'arbre de décision sert à la production des règles de classification d'individus ou d'objets. Ces règles sont en général de type « SI – ALORS ». Ceci peut se faire à travers l'ensemble des chemins partant de la racine de l'arbre à chacune des feuilles. Chaque chemin représente une règle de classification.

La méthode CART (Classification And Regression Tree) constitue l'un des types de construction d'arbre les plus connus. Cette méthode utilise une recherche de grille exhaustive de toutes les segmentations univariées possibles en vue de trouver la segmentation optimale d'un arbre de décision.

Les arbres de décision présentent plusieurs avantages. On peut citer, la prise en compte simultanée de variables quantitatives et qualitatives, la sélection des variables les plus informatives, et la production de règles logiques de classification aisément interprétables. Comme ils sont considérés comme un outil incontournable pour l'extraction des connaissances et du data mining en général.

Cependant, ces arbres présentent un certain nombre de limites. Primo, dans le traitement des variables numériques, le choix des seuils ne prend pas en considération les propriétés de densité (proximité) des valeurs.

Secondo, ils sont sensibles aux variations, ce qui entraîne des variations dans les arbres produits et de leurs performances.

La théorie des ensembles approximatifs

En 1982, dans un article fondateur de la revue « International Journal of Information and Computer Sciences », Pawlak ¹¹ a introduit la théorie des ensembles approximatifs (TEA). Cette théorie vise l'analyse et la classification des données compte tenu de leur imprécision. Elle présente l'avantage d'être indépendante de la distribution des données. Plusieurs domaines ont réussi l'application de la TEA. Parmi eux, le domaine médical, la finance, le marketing, ... etc. La TEA pourrait constituer une méthode adéquate pour la classification en GRH du fait de la nature non paramétrique des données collectées ainsi que leur imprécision.

Sur le plan informatique, il existe aujourd'hui deux logiciels très conviviaux, téléchargeables sur Internet, qui permettent, en fait, l'application de la TEA. Il s'agit des logiciels ROSE2 et ROSETTA.

a) Terminologie de la TEA

La TEA présente les avantages suivants :

- Le raisonnement à partir des données imprécise ou ambiguë ;
- L'utilité pour l'apprentissage automatique et l'extraction de connaissances ;

-

¹¹ Pawlak Z. 1982. Rough Sets. *International Journal of Information and Computer Sciences*. Vol 11. pp. 341-356.

- La réduction du degré de précision en vue de définir le concept afin de mettre en relief des relations.

Cette théorie vise à prévoir l'affectation des individus ou des objets à plusieurs ensembles afin de gérer l'indiscernabilité.

Pour ce faire, les données collectées se présentent sous la forme d'une table d'information qui est une matrice à deux dimensions divisées en trois sections dont chaque ligne décrit une entité. Ces sections sont :

- Les éléments de l'univers : les individus ou les objets concernés ;
- Les conditions : reflètent les caractéristiques des éléments de l'univers à étudier ;
- Les décisions : représentent les caractéristiques que l'on cherche à expliquer, à prédire ou à classer.

L'application de la TEA passe par plusieurs phases : discrétisation, formation des atomes, recherche des redondances, génération des règles, classification et validation.

b) Discrétisation des données

L'application des principes des ensembles approximatifs exige que toutes les données (appelées aussi attributs) soient sous forme discrète. Donc les données continues doivent être discrétisées avant de commencer l'opération de génération de règles.

La discrétisation granule le domaine continu des attributs, ce qui facilite l'induction de règles courtes et fortes. Il existe plusieurs manières de discrétiser des données continues. La première consiste en la création des classes sur la base des connaissances de la personne. Cependant, il n'est pas toujours facile d'établir un tel classement. La deuxième se traduit par le recours à certains algorithmes permettant la transformation des données continues en données discrètes. On pourrait citer, à titre d'exemple, le logiciel ROSE2 qui dispose de trois types de discrétisation :

- Locale
- Locale supervisée
- Globale

c) Formation des atomes

La TEA repose sur un concept fondamental qui est la relation d'indiscernabilité, normalement associée à un ensemble d'attributs. Or, les ensembles indiscernables sont appelés ensembles élémentaires. Lorsque tous les attributs sont pris en compte, les ensembles élémentaires cités ci-dessus formés par les objets sont appelés atomes. De plus, les attributs décisionnels peuvent être exprimés de la même façon. Ils sont appelés donc concepts.

Sur la base des atomes, on peut évaluer comment les différents individus ou objets se placent par rapport aux concepts.

En ce qui concerne la performance de l'application, deux mesures sont connues : l'exactitude et la fiabilité.

L'exactitude est une mesure de l'inconsistance des données. Elle est donnée par l'exactitude de l'approximation calculée comme l'approximation inférieure sur l'approximation supérieure d'un concept. La deuxième mesure, qui est la fiabilité permet d'apprécier la qualité de l'approximation des données. Cette qualité est définie comme la somme des individus ou objets des approximations inférieures pour tous les concepts sur le nombre d'individus ou objets totaux.

d) Recherche des redondances

Il est aisé de définir les attributs redondants par le concept d'indiscernabilité. En effet, si un ensemble d'attributs et un sous-ensemble d'attributs définissent la même relation d'indiscernabilité, alors chaque attribut appartenant au sous-ensemble est redondant.

e) Génération de règles

Les règles générées sont de type « Si – Alors » sur la base des atomes et concepts formés précédemment. Il y a des règles certaines (induites par l'approximation inférieures) et des règles possibles (induites par l'approximation supérieure du concept).

f) Classification

Une fois les règles formulées, la classification pourra s'effectuer. Si la méthode discrimine bien entre les segments, la classification sera bonne et vice-versa. Dans la pratique, il existe deux types de classification : a priori et post-hoc.

RÉSULTATS ET DISCUSSION

Cette étude présente un modèle composé de sept (07) variables. Les six premiers représentent des attributs caractérisant le dirigeant public. La septième variable constitue le segment auquel ce dirigeant appartient.

Toutes ces variables sont discrètes (catégorielles). Dans le but de réaliser l'objectif de cette étude, le réseau de neurones a été implanté sous Excel et l'arbre de décision sous Statistica 5.5 (StatSoft, 1997)¹². Par ailleurs, le logiciel Automatic Problem Solver (APS) 3.0 (2004) permet de réaliser une classification en utilisant la technique des algorithmes génétiques. Enfin, la théorie des ensembles approximatifs trouve son application à travers le logiciel ROSETTA¹³. L'objectif de cette étape est d'obtenir les résultats de l'application de chacune de ces techniques de segmentation, puis les comparer entre eux afin de dégager l'outil adéquat qui sera capable de reconnaître la structure du modèle de segmentation de la population des dirigeants publics. Ainsi, notre échantillon a été divisé en deux sous-

¹² Statsoft. 1997. Statistica for windows. Version 5.5. France.

Ohrn A. and Komorowski J. 2001. ROSETTA: A Rough Set Toolkit for Analysis of Data, http://www.idi.ntnu.no/~aleks/rosetta/

échantillons à raison de 70% et 30%. Le premier constitue celui d'apprentissage, et le second servira pour le test de la structure de segmentation résultante.

Résultats du réseau neuronal

Cet outil est considéré comme la méthode d'apprentissage par excellence. Il s'agit d'introduire les six attributs au réseau, et laisser la boite noire se paramétrer elle-même, en espérant une erreur de classification la moins grande possible.

Dans le cadre de cette étude, le réseau de neurones utilisé est un réseau de type perceptron multicouche constitué de 4 couches. La couche d'entrée comporte 6 neurones. Les deux couches cachées (3 et 4 neurones respectivement) servent pour le calcul et la propagation de l'erreur. Enfin, la couche de sortie représente l'output désiré, en d'autres termes, le segment auquel l'individu est affecté. Le taux d'apprentissage est fixé à 0.1. Ainsi que le nombre d'itérations se limite à 500 itérations. L'algorithme d'apprentissage est, bien entendu, la rétro-propagation.

Après exécution de cet algorithme sous Excel, les résultats de la classification montrent que les observations ayant été correctement classées représentent 66.67 %, et celles relevant du mauvais classement représentent 23.08%.

Résultats de l'algorithme génétique

Un algorithme génétique de classification a été quant à lui, implanté sous APS 3.0 (2004). La procédure d'apprentissage utilise les mécanismes de base de l'algorithme génétique pour faire évoluer et optimiser la solution finale, à savoir la mutation et le croisement. Un processus de sélection des meilleurs chromosomes à reproduire et de ceux à remplacer s'ajoute à ces deux opérations. Les paramètres considérés à cette fin sont les suivants :

Nombre de chromosomes : 100 ; Nombre de générations : 200 ; Nombre de gènes : 5 ; Taux de mutation : 0.05 ; Taux de croisement : 0.1 :

Fonction fitness: Maximum Fitness.

Selon cette méthode, le pourcentage du bon classement est de l'ordre de 60.87 %. Par contre, les observations mal classées sont estimées à 39.13 %. On s'aperçoit que l'algorithme génétique construit n'a pas réalisé la performance souhaitée.

Résultats de l'application de l'arbre de décision

Un algorithme de construction d'arbre de décision de type Recherche Exhaustive de Segmentations Univariées Style – CART (déviance fixée à 5) montre un arbre relativement simple comprenant 3 segments et 4 nœuds terminaux. Une validation croisée effectuée 15 fois montre que l'erreur du mauvais classement se situe autour de 46,8%, soit 53,2% de bon classement.

De plus, il ressort de l'analyse que les prédicteurs importants sont : l'expérience du dirigeant dans la fonction ou le poste (100%), sa fonction actuelle (71%), niveau d'instruction (53%), son âge (48%), le type d'administration qui l'emploie (40%), et enfin son genre (9%). On en déduit que la variable « sexe » ne participe pas à la définition des segments de dirigeants. En conclusion, cet arbre peut générer trois 4 règles de classification.

Résultats de l'application de la TEA

Les attributs de la présente étude ne nécessitent pas une discrétisation du fait qu'ils sont de type discret (catégoriel). Le logiciel ROSETTA (2001) a été utilisé pour classer les dirigeants en trois segments différents suivant leurs caractéristiques socioprofessionnelles (six attributs). Après différents essais, il s'est avéré que la variable « sexe » du dirigeant n'exerce aucune influence sur la segmentation, et n'améliore pas suffisamment la classification du fait que l'échantillon est composé d'hommes à 86%. Par conséquent, elle ne sert pas à différencier les dirigeants interrogés. En d'autres termes, on peut dire que les hommes et les femmes ont le même sens du travail, style de commandement, et les mêmes préoccupations de GRH. Donc, dans un souci de parcimonie, l'attribut « sexe » peut être supprimé.

Ainsi, le pourcentage des observations correctement classées s'élève à 83%. Par contre, le mauvais classement constitue 17%.

Par ailleurs, le segment pour lequel la segmentation est la meilleure est le segment N° 1 avec 92,6%. Notons qu'aucun segment n'atteint 100%. Les deux autres segments ont obtenu une bonne classification (81,5% et 73,7% respectivement).

En résumé, la classification issue de l'application de la TEA est très bonne (83%) par rapport aux autres techniques appliquées.

Discussion

Au vu de cette étude, la TEA a enregistré un très bon taux de segmentation. De plus, ce résultat n'est pas surprenant du moment où cette théorie présente plusieurs avantages lors de sa mise en œuvre. En premier lieu, son application ne dépend pas de la distribution des données collectées, et qui sont généralement de type nominal.

En deuxième lieu, cette méthode permet d'inclure des données subjectives puisqu'elle travaille avec des classes et non avec des valeurs numériques (chiffres). Enfin, elle permet de générer un ensemble de règles de classification caractérisant les différents individus dans le but d'aider le décideur en matière de détection et de sélection des futurs candidats aux fonctions importantes de l'administration publique à travers leurs attributs.

A cet égard, le même logiciel utilisé précédemment, permet d'obtenir des règles de segmentation par le biais d'un algorithme spécifique. Au total, 46 règles de classification ont été apprises : 35 sont exactes et 11 approximatives. Le premier segment est décrit par 12 règles exactes, le deuxième par 14 règles et le dernier par 9 règles. Notons que chacun des trois segments a été décrit aussi par six (06) règles approximatives. Ces règles se lisent : «SI....et... etALORS..... ». Elles peuvent désormais s'énoncer en « langage naturel » comme suit :

« SI un dirigeant ayant un niveau universitaire ET il est directeur ET travaille dans une administration centrale ALORS il est probable qu'il appartienne au segment N°3. Donc il serait convenable de lui appliquer le profil qui correspond à ce segment. Les règles obtenues (output de ROSETTA) sont les suivantes (le chiffre entre deux parenthèses indique le code de la modalité) :

Règles exactes:

```
- scolarite(1) AND experience(2) => segment(1)
```

- scolarité(1) AND expérience (7) => segment (1)
- scolarite(1) AND experience(6) => segment(1)
- scolarite(4) => segment(1)
- age(4) AND experience(2) => segment(1)
- age(2) AND scolarite(1) => segment(1)
- scolarite(1) AND admin(1) => segment(1)
- experience(7) AND fonction(1) => segment(1)
- age(2) AND admin(1) => segment(1)
- age(2) AND experience(2) AND fonction(3) => segment(1)
- admin(4) => segment(1)
- experience(3) AND fonction(5) AND admin(1) => segment(1)
- age(3) AND scolarite(2) AND fonction(5) => segment(2)
- Scolarité (3) AND expérience (4) => segment(2)
- age(2) AND experience(3) AND fonction(4) => segment(2)
- age(1) AND experience(4) AND fonction(4) => segment(2)
- experience(1) => segment(2)
- age(1) AND experience(2) => segment(2)
- age(3) AND experience(5) => segment(2)
- age(4) AND experience(6) => segment(2)
- age(1) AND scolarite(3) => segment(2)
- scolarite(2) AND admin(2) => segment(2)
- scolarite(3) AND admin(1) => segment(2)
- experience(2) AND fonction(4) => segment(2)
- experience(4) AND fonction(3) => segment(2)
- experience(7) AND fonction(5) => segment(2)
- age(2) AND experience(5) => segment(3)
- scolarite(3) AND experience(3) => segment(3)
- scolarite(3) AND experience(7) => segment(3)
- scolarite(1) AND experience(5) => segment(3)

- age(1) AND admin(3) => segment(3)
- scolarite(3) AND fonction(3) => segment(3)
- -scolarite(2) AND fonction(4) AND admin(1) => segment(3)
- age(1) AND experience(3) AND fonction(4) => segment(3)
- age(3) AND fonction(1) => segment(3)

Règles approximatives

- Age(4) AND scolarité (1) AND fonction(3) => segment(1) OR segment(3)
- age(2) AND experience(3) AND fonction(3) => segment(1) OR segment(2)
- age(2) AND experience(4) AND fonction(4) => segment(1) OR segment(2)
- age(1) AND experience(4) AND fonction(5) => segment(2) OR segment(3)
- age(3) AND scolarite(1) AND experience(3) AND admin(4) => segment(2) OR segment(1)
- age(3) AND scolarite(2) AND experience(2) AND fonction(3) => segment(1) OR segment(3)
- age(2) AND scolarite(2) AND experience(3) AND fonction(5) => segment(1) OR segment(3)
- age(3) AND scolarite(2) AND experience(3) => segment(3) OR segment(1)
- age(2) AND scolarite(2) AND fonction(5) AND admin(4) => segment(1) OR segment(3)
- age(1) AND experience(3) AND fonction(3) AND
- scolarite(2) AND experience(6) AND fonction(3) => segment(2) OR segment(1)
- age(4) AND fonction(1) => segment(2) OR segment(1)

Ensuite, ces règles peuvent être exploitées en vue de décrire le profil des segments. Pour ce faire, les règles qui concernent un même segment ont été placées côte à côte. Cela facilitera la caractérisation des segments.

Segment N° 1

Les règles dont le segment est numéroté 1 caractérisent ce groupe de dirigeants. Cette catégorie regroupe des dirigeants disant jeunes (entre 31 et 39 ans dans la majorité des cas). Ces derniers possèdent un niveau secondaire ou un autre diplôme professionnel. Leurs expériences varient de 5 à 16 ans. Ils occupent, actuellement des postes de chefs de service, de secrétaires généraux, et de directeurs généraux (effet de l'ancienneté).

Enfin, ces dirigeants travaillent principalement dans des administrations centrales (départements ministériels) et dans des établissements publics administratifs.

Segment N° 2

Concernant ce segment, il est à noter que les deux attributs «Âge » et « Expérience » ne participent pas à la discrimination des dirigeants. Ce segment regroupe les dirigeants qui disposent d'un niveau un universitaire et/ou post-universitaire. Ils occupent actuellement des fonctions de directeurs, de sous-directeurs et de chefs de services. Bien entendu, les administrations qui les emploient sont la centrale et ses services déconcentrées. En réalité, ce résultat est très justifié car les services centraux recourent généralement au produit de formation délivré par les grandes écoles et les meilleurs établissements de l'enseignement supérieur (ENA, École Polytechnique, ... etc.). Les administrations déconcentrées profitent ainsi de ces recrutements et promotions.

Segment N° 3

Ce dernier segment est caractérisé par plusieurs règles de classification. On peut noter aussi que les deux attributs « Fonction » et « Administration » ne participent vraiment à la segmentation des dirigeants. Ce segment regroupe les dirigeants ayant généralement le rang de chef de service, possédant une expérience de 3 à 15 ans. Ils possèdent également un niveau secondaire ou un niveau universitaire (la plupart des cas). Ils occupent des postes dans des collectivités locales (communes et *wilayates*).

Enfin, on peut dire qu'il existe trois segments représentant la population des dirigeants de l'administration publique algérienne à travers l'échantillon de cette étude. Cette segmentation reflète dans une large mesure la culture administrative et les pratiques managériales qui règnent au sein de la fonction publique algérienne.

Synthèse et perspectives

Cette étude visait la segmentation de la population des dirigeants de l'administration publique à travers leurs attributs. Différentes techniques empruntées à l'intelligence artificielle ont été utilisées à cette fin. Le but étant d'apprendre des règles de classification permettant d'aider l'autorité de nomination de connaître le profil du candidat à la fonction ou au poste supérieur.

Ceci dit, les dirigeants se distinguent bien les uns des autres à travers l'utilisation de la théorie des ensembles approximatifs (TEA). Cette théorie présente une multitude d'avantages. Premièrement, sa mise en œuvre ne dépend pas de la distribution des données. Elle permet également, de prendre en compte l'incertitude et l'imprécision des données collectées au travers de la création de classes. Enfin, cette méthode offre un ensemble de règles permettant la discrimination et la classification des individus tout en étant aisément interprétables.

Ainsi, cette étude montre les apports, dans le domaine de la GRH, de l'approche qui repose sur l'utilisation d'un ensemble de techniques appelé « Soft Computing » pour apprendre des règles floues de classification¹⁴ (Shapiro, 2000). Ces règles ainsi apprises servent non seulement à prédire le segment auquel appartient un nouveau candidat mais aussi à mieux

Shapiro A. F. 2000. Soft Computing Applications in Actuarial Science. Working paper. Penn State University.

connaître et analyser la population des dirigeants en identifiant les variables qui décrivent mieux les segments obtenus.

Cependant, cette étude présente plusieurs limites ou points à améliorer pour que la structure obtenue soit opérationnelle. En premier lieu, la taille de l'échantillon étudié s'avère insuffisante. Son augmentation serait d'une importance cruciale.

En deuxième lieu, il est intéressant d'intégrer de nouvelles variables telles que le statut familial du dirigeant, son revenu, le nombre de perfectionnements et de formations, ... etc. Ces variables pourraient améliorer la performance de la segmentation.

En dernier lieu, le nombre important de règles générées (46 règles) rend l'opération d'affectation et de sélection plus difficile. Il est conseillé de réduire au maximum ce nombre tout en garantissant une bonne parcimonie de la structure de classification. Parmi les pistes prometteuses, on peut citer l'approche par les ensembles approximatifs basée sur la dominance (Dominance – Based Rough Set Approach). Cette technique combine les avantages de la TEA et la relation de dominance ¹⁵ (Greco et al, 2002). En effet, cette approche permettra d'obtenir une classification multicritère rénovée.

REMERCIEMENTS

Je tiens à remercier chaleureusement M. Kissouri (cabinet MBS d'Alger), M. Ramdhani (ENSP) et M. Essaid Taïb, professeur à l'ENA d'Alger qui ont apporté une aide précieuse pour l'administration du questionnaire de cette étude.

RÉFÉRENCES

Bolgot, S. et Meyfredi, J. C. (1999). Réseaux de neurones, lissage de la fonction d'actualisation et prévision des OAT démembrées : une étude empirique. Document GREQAM, Université de la Méditerranée, France.

Collins, J. M. and Clarck, M. R. (1993). 'An application of the theory of neural computation to the prediction of workplace behaviour: an illustration and assessment of network analysis'. Personnel Psychology, 46: 503-524.

Cordon, O., Gomide, F., Herrera, F., Hoffmann, and Magdalena, L. (2004). 'Ten Years of Genetic Fuzzy Systems: Current framework and News Trends'. Fuzzy Sets and Systems, 141: 5-31.

Greco, S., Matarazzo, B. and Slowinski, R. (2002). 'Rough Approximation by Dominane Relations'. International Journal of Intelligent Systems, 17(2): 153-171.

Liu, W., Lepak, D. P., Takeuchi, R. and Sims, H. R. (2003). 'Matching leadership styles with employment modes: strategic human resource management perspective'. Human Resource Management Review, 13: 127–152.

¹⁵ Greco S., Matarazzo B. and Slowinski R. 2002. Rough Approximation by Dominane Relations. International Journal of Intelligent Systems, 17(2): 153-171.

Montagno, R., Sexton, R. S. and Smith, B. N. (2002). Using neural networks for identifying organizational improvment strategies, www.faculty.smsu.edu/r/rss000f.

Nguyen Phuong, T., Cliquet, G. et Leroy, F. (2003). 'L'opposition entre la taille du marché et degré d'homogénéité des segments : approche par la logique floue'. Décision Marketing, 32: 55-69.

Ohrn, A. and Komorowski, J. (2001). ROSETTA: A Rough Set Toolkit for Analysis of Data, http://www.idi.ntnu.no/~aleks/rosetta/

Paquet, P. (1997). L'utilisation des réseaux de neurones artificiels en finance. Document de recherche n° 1. Laboratoire Orléanais de Gestion.

Pawlak, Z. (1982). 'Rough Sets'. International Journal of Information and Computer Sciences, 11: 341-356.

Shapiro, A. F. (2000). Soft Computing Applications in Actuarial Science. Working paper, Penn State University.

Statsoft. (1997). Statistica for windows. Version 5.5. France.

Taïb, E. (1996). 'La Réforme de l'administration en Algérie'. Annuaire de l'Afrique du Nord, XXXV : 343-379.