

Traiter simultanément des variables qualitatives et quantitatives à partir de modèles non linéaires

Application des CAO à une typologie des
modes d'organisation du travail

Patrick Rousset

Céreq,
10, place de la Joliette, BP 21321, 13567 Marseille cedex 02.
E-mail : rousset@cereq.fr

Jean-François Giret

Iredu - Université de Bourgogne
E-mail : Jean-François.giret@u-bourgogne.fr

SOMMAIRE

SOMMAIRE	1
INTRODUCTION	3
1. MÉTHODE	3
2. UN EXEMPLE DE DONNÉES SIMULÉES	4
3. APPLICATION DANS LE CADRE ÉCONOMIQUE : LES MODES D'ORGANISATION DU TRAVAIL	8
3.1. L'approche économique des modes d'organisation du travail	8
3.2. La typologie des modes d'organisation du travail	8
4. CONCLUSION	11
BIBLIOGRAPHIE	12

INTRODUCTION

Le traitement simultané des variables qualitatives et quantitatives est souvent nécessaire dans l'analyse des données. L'usage dans ce cas, consiste à transformer en qualitatif toute l'information ce qui revient dans la pratique à en perdre une grande partie pouvant servir à mesurer les distances entre individus ou variables. D'autres codages ont déjà été proposés pour éviter ce problème dans le cadre de l'analyse linéaire. Mais la catégorisation répond aussi souvent au besoin d'introduire du non-linéaire dans les méthodes linéaires ce qui a réduit la diffusion de tels codages. Au contraire, lorsqu'on utilise des modèles non linéaires pour optimiser l'ajustement, il est contreproductif qu'une partie de cet ajustement soit due à un pré-codage arbitraire très réducteur d'information. L'objet de cet article est donc de reconsidérer l'intérêt de codages spécifiques au traitement simultané quantitatif-qualitatif dans le cadre non linéaire. En particulier, on s'intéressera à un procédé de pré-codage, proposé par Brigitte Escoffier-Cordier [5] qui, en étant couplé à la distance du χ^2 , conserve les propriétés des grandeurs quantitatives. Les cartes d'auto-organisation, qui offrent un modèle de classification non supervisé, sont un bon support d'illustration. En effet, la propriété principale des CAO en analyse de données est la conservation de la topologie : La proximité des classes sur la carte de sortie rend compte de celle de leurs individus dans l'espace des données. L'efficacité de cette propriété dépend donc fortement de la capacité des variables à intégrer la proximité des individus en fonction d'une mesure (qui est donc réduite par le traitement des données qualitatives et quantitatives). De plus, les CAO fournissent une représentation graphique de la prise en compte des proximités qui constitue un outil visuel adapté pour illustrer notre propos. Nous présentons ainsi dans ce papier l'apport aux CAO d'un procédé de pré-codage, couplé à la distance du χ^2 , qui conserve les propriétés des grandeurs quantitatives. Nous utiliserons dans un premier temps, pour illustrer son apport, des données simulées puis nous l'appliquerons, dans le cadre économique, à des données de l'enquête « Génération 98 », sur les modes d'organisation du travail.

1. MÉTHODE

La pratique la plus commune de traitement de l'information, dans le cadre d'un mélange de données qualitatives et quantitatives, consiste à recomposer les variables en code binaire (comme dans le tableau de Burt). Sous cette forme, il est alors possible d'appliquer des méthodes conçues pour les données qualitatives comme par exemple l'analyse des correspondances [4] ou, pour les cartes d'auto-organisation, le KACM [1]. Le recodage qualitatif d'une variable quantitative V consiste à diviser l'intervalle où elle prend ses valeurs en plusieurs sous intervalles, et ensuite de recueillir, dans les modalités d'une variable qualitative V' équivalente, l'intervalle d'appartenance de l'individu. L'information est alors dispatchée dans autant de variables que d'items, chacune correspondant à une indicatrice de présence de l'individu dans l'intervalle associé à l'item. La signification du codage de l'item (donc de la mesure) étant ignorée, ce procédé fait perdre la majeure partie de l'information en termes de proximité d'individus : la seule information conservée est leur éventuelle appartenance au même intervalle alors que l'appartenance à des intervalles contigus ou proches est, elle, perdue. Au final, quand on considère l'ensemble des données, la proximité entre deux individus dépend uniquement du nombre de modalités communes mais plus de la proximité entre ces modalités. D'un autre côté, les cartes d'auto-organisation ([3], [2]) sont extrêmement liées à la notion de degré de proximité, à la fois pour leur interprétation (la lecture de la carte associée) et pour la phase d'apprentissage de l'algorithme. Les cartes d'auto-organisation sont communément considérées comme une généralisation de l'algorithme des centres mobiles, en particulier dans sa version stochastique proposée par

Mac Queen [9]. A chaque itération, l'information contenue par l'individu sert dans un cas (centres mobiles) à mettre à jour le centre de classe le plus proche alors que dans l'autre sont également mis à jour les voisins, définis par une fonction de voisinage qui décroît avec le temps. Il est donc essentiel pour le comportement de l'algorithme que la notion de voisinage prenne au mieux en compte la proximité réelle (mesurée et non recodée) des individus.

Pour éviter ces inconvénients, Brigitte Escoffier-Cordier [5] propose une transformation des variables quantitatives en profils qui permet d'appliquer l'analyse des correspondances simultanément à des variables quantitatives V et qualitatives V' . Il consiste à décomposer chaque variable quantitative V centrée normée de valeur x en deux variables V_1 et V_2 de valeurs $\frac{1}{2}-x$ et $\frac{1}{2}+x$. On peut remarquer que pour l'individu i , la somme V_1+V_2 vaut 1 et que la somme en colonnes de V_1 et V_2 valent $n/2$. Enfin, dans le cadre de l'analyse en correspondance, la recherche de facteurs principaux relève du même algorithme pour l'espace engendré par l'ensemble des couples (V_1, V_2) que pour l'ensemble des variables qualitatives et donc que l'ensemble constitué des variables de type V_1, V_2 et V' .

Du point de vue des cartes d'auto-organisation, on peut remarquer également que la distance naturelle pour des profils comme pour des données qualitatives est celle du χ^2 et donc les cartes d'auto-organisation peuvent aussi être appliquées aux données codées comme précédemment.

2. UN EXEMPLE DE DONNÉES SIMULÉES

A titre d'exemple, nous avons appliqué la méthode à deux cas de la même distribution : le cas où l'on dispose de la valeur quantitative brute et le cas où l'on subit une préalable catégorisation en variable qualitative ordinale et où on décide de reconstituer l'information quantitative en considérant la signification des items. Cette dernière situation, très fréquente, résulte de l'usage qui consiste définir une enquête dès la conception comme qualitative ou quantitative : la première option aboutissant à la catégorisation des questions/réponses concernant des dimensions quantitatives. A fin de comparaisons, la méthode est comparée à l'approche classique appliquée aux variables qualitatives (de même catégorisation que le deuxième cas précédent mais cette fois en ignorant la valeur des items). Nous avons simulé une distribution uniforme simulée de 1 000 individus sur le carré $[0,5[x[0,5[$ et considéré les variables associées *abscisse* et *ordonnée*. Nous avons dans un premier cas conservé la valeur dite quantitative, dans un deuxième catégorisé les abscisses et ordonnées en codant de 0 à 4 leur intervalle unitaire d'appartenance (ce qui est équivalent à la partie entière de la valeur quantitative). Nous avons appliqué la méthode proposée à ces deux cas. La méthode de référence, dite classique, a consisté à reprendre le deuxième cas en ignorant la valeur des items par le processus habituel revenant à dichotomiser l'appartenance à chacun des intervalles unitaire. Les trois approches de ces données ont été traitées par les cartes d'auto-organisations avec la distance du χ^2 . On attend bien sûr des cartes résultantes de ces approches qu'elles respectent la conservation de la topologie par rapport aux valeurs brutes. C'est à dire, dans notre exemple, qu'elles rendent compte visuellement de la distribution uniforme sur le plan. Les approches 1 et 2 qui sont bien sur plus naturelles dans cet exemple répondent de façon satisfaisante au problème. De son côté, la méthode 3, qui ignore la structure interne des variables, est donc, de fait, vouée à un échec lorsque les grandeurs sont indépendantes comme ici. Le résultat illustre très bien le problème de ce codage car la structure de voisinage ne rend pas compte de l'ordre sur l'échelle des abscisses ou ordonnées et peut donc regrouper, par exemple, deux individus d'abscisses éloignées mais de coordonnées semblables.

Pour illustrer les résultats, les unités de la carte sont autant de fenêtres graphiques dans lesquelles est représentée la distribution de la classe associée. Un premier type de graphique (Tableau 1) pointe les individus de la classe sur le plan $[0,5[x[0,5[$. Un deuxième type (Tableau 2) représente par un camembert la répartition du contingent de la classe sur les cinq items de l'abscisse et de l'ordonnées. Dans le cas de la méthode proposée appliquée au codage quantitatif brute (tableaux 1 et 2, ligne 1), la continuité de la distribution et la proximité des classes traduisent bien celle de l'espace des données. Dans l'approche

qualitative ordinale (tableaux 1 et 2, ligne 2), on retrouve les mêmes conclusions si ce n'est que ce modèle traduit moins bien la continuité de l'espace des données. L'approche qualitative classique (tableaux 1 et 2, ligne 3) montre bien la perte d'information lors du prétraitement de l'information. Dans cette approche, seuls trois niveaux de distance entre les points sont pris en compte : 0 (abscisses et ordonnées de même item), 1 (abscisses ou ordonnées de même item) et 2 (abscisses et ordonnées d'items différents). Par conséquent, les points (2, 3) et (4, 3) sont à la même distance (de valeur 1) du point (0, 3) que les point (1, 3), (0, 4) et (0, 2). L'unité 1 de la carte du tableaux 2 ligne 3 traduit cette propriété de l'approche purement qualitative. Elle regroupe en effet des zones de l'espace des données brutes éloignées et non connexes : par exemple les points des zones $[0,1[\times]3,4[$, $[2,3[\times]3,4[$ et $[4,5[\times]3,4[$.

Enfin, un troisième type de graphique (Tableau 3) représente le croisement de la classification avec une variable exogène, choisie fortement corrélée. Elle vaut 0 si l'abscisse est inférieure à 2,5 et 1 sinon. Cette variable divise en deux zones l'espace des données. Cette variable étant très corrélée avec les données brutes, on attend qu'elle soit très bien expliquée par la carte. A l'arrivée, avec la méthode proposée, pour les deux codages, il est très simple de retrouver les deux zones sur la cartes (tableau 3, lignes 1 et 2) alors que c'est impossible avec la méthode classique (tableau 3, ligne 3).

Tableau 1

Répartition sur une CAO d'une distribution uniforme sur le carré $[0,5[\times]0,5[$. Dans chaque unité de la carte, les individus de la classe associée sont pointés sur le plan $[0,5[\times]0,5[$.

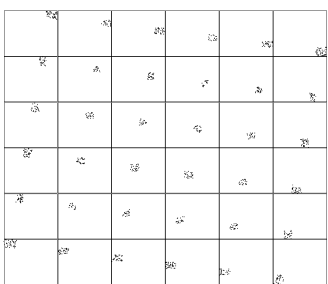
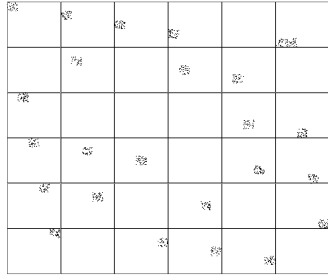
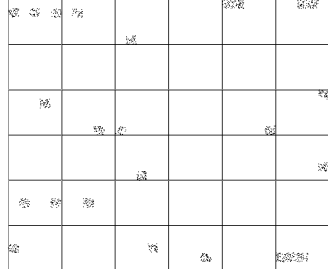
<p>Cas 1 - méthode de l'article :</p> <ul style="list-style-type: none"> - le codage de l'abscisse et de l'ordonnée est transformé et reste quantitatif. - L'analyse est de type qualitative. <p>Les zones de projections sont connexes et la proximité entre les classes sur la carte coïncide à une proximité sur le plan. La continuité la distribution est de plus bien traduite.</p>	
<p>Cas 2 - méthode de l'article :</p> <ul style="list-style-type: none"> - Le codage qualitatif qui renseigne le numéro de l'intervalle unitaire de l'abscisse et de l'ordonnée est transformé et devient quantitatif. - L'analyse est de type qualitative. <p>Les zones de projections ne sont pas connexes et la proximité entre les classes sur la carte coïncide mal avec une proximité sur le plan.</p>	
<p>Cas 2 - méthode qualitative classique :</p> <ul style="list-style-type: none"> - Le codage qualitatif renseigne le numéro de l'intervalle unitaire et traduit en variables qualitatives binaires en faisant perdre la valeur des items. - L'analyse est de type qualitative. <p>Les zones de projections ne sont pas connexes et la proximité entre les classes sur la carte coïncide mal avec une proximité sur le plan.</p>	

Tableau 2

Répartition des variables endogènes (abscisses et ordonnées) dans les classes.

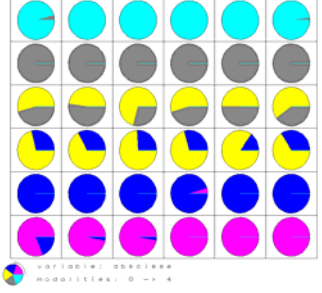
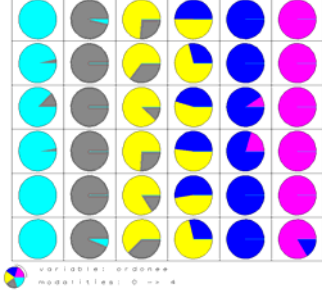
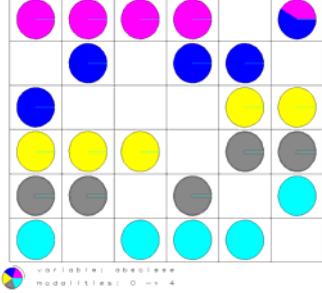
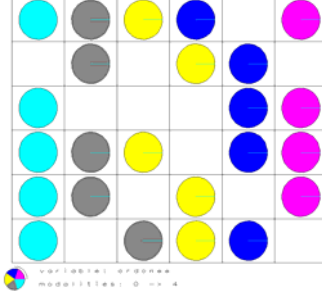
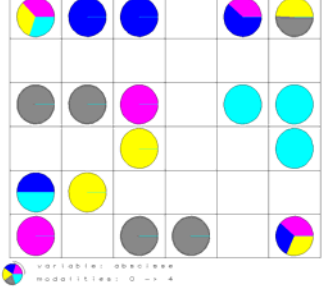
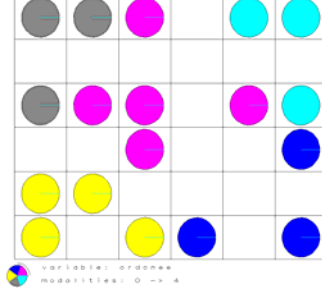
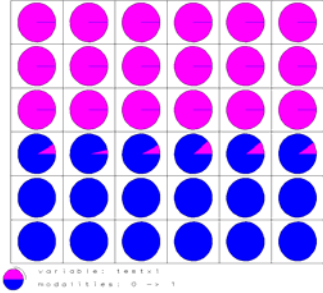
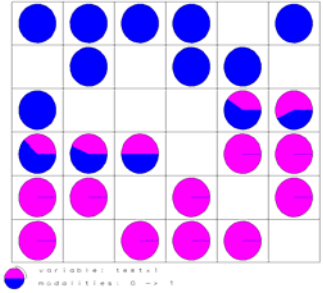
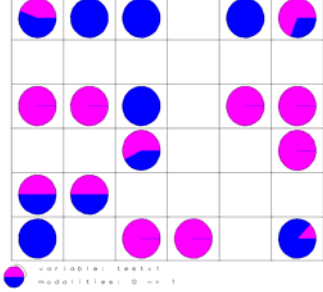
Méthode	abscisses	ordonnées
<p>Cas 1 - méthode de l'article : La structure de carte rend compte de la structure de la distribution. Les zones de projections sont connexes et la proximité entre les classes sur la carte coïncide à une proximité sur le plan.</p>		
<p>Cas 2 - méthode de l'article : idem méthode 1</p>		
<p>Cas 2 - méthode qualitative classique : La structure de carte ne rend pas compte de la structure de la distribution. L'unité 1 mélange les abscisses 0, 2 et 4</p>		

Tableau 3

Croisement des classes avec une variable exogène fortement corrélée : l'indicatrice pour l'abscisse d'appartenance à l'intervalle [0,2.5[.

<p>Cas 1 - méthode de l'article : La corrélation de cette variable avec les variables endogènes est bien rendue.</p>	
<p>Cas 2 - méthode de l'article : idem méthode 1</p>	
<p>Cas 2 - méthode qualitative classique : l'indicatrice pour l'abscisse d'appartenance à l'intervalle [0,2.5[. La distribution de cette variable est très mal interprétée par la carte. Les modalités 0 ou 1 se retrouvent un peu partout sur la carte.</p>	

3. APPLICATION DANS LE CADRE ÉCONOMIQUE : LES MODES D'ORGANISATION DU TRAVAIL

3.1. L'approche économique des modes d'organisation du travail

Nous proposons dans ce travail d'apporter un éclairage empirique sur les modes d'organisation du travail dans lesquels sont insérés les jeunes. Plus précisément, il s'agit d'identifier les différents types d'organisation du travail, à partir d'informations sur l'environnement de travail des jeunes en début de carrière professionnelle. Nous utilisons un module spécifique de l'enquête « Génération 98 » du Cereq sur les sept premières années de vie active. L'enquête « Génération 98 » permet d'analyser les cheminements professionnels d'une cohorte de jeunes sortis de formation initiale en 1998 et interviewés en 2001, 2003, 2005. L'échantillon interrogé, environ 16 000 jeunes en 2005, est représentatif des 750.000 individus jeunes sortis du système éducatif cette année-là. Le module « expérience et organisation du travail » comporte notamment des informations sur les caractéristiques de l'organisation du travail où se trouve les jeunes, les relations avec la hiérarchie et les collègues de travail, le type de savoir que les jeunes doivent acquérir (très codifié ou non). Il est également possible d'examiner les liens entre ces différents modes d'organisation du travail et la stabilisation professionnelle des jeunes après sept années de vie active.

De nombreux travaux ont analysé les transformations des modes d'organisation du travail, pointant pour certains l'émergence de nouveaux modèles, répondant mieux aux contraintes économiques actuelles et aux modes d'incertitude qui pèsent sur les entreprises. Ainsi Lorenz et Valeyre ([6] et [7]) proposent une typologie en 4 classes de ces modes d'organisation du travail dans les pays européens à partir des conditions de travail des salariés : classe 1 (39% des salariés) - les organisations apprenantes, caractérisées par une forte autonomie et un contenu du travail très cognitif, classe 2 (28%) - les organisations en « lean production », où les contraintes de travail sont lourdes malgré des pratiques de travail en équipe et un contenu du travail fortement cognitif mais avec une autonomie restreinte-, classe 3 (14%) - les organisations tayloriennes, avec des tâches répétitives et peu de marge d'autonomie- et Classe 4 (19%) - Les organisations de structure simple, sans organisation structurée et avec un contrôle direct du supérieur hiérarchique - .

Nous nous proposons dans ce travail de reprendre un ensemble de variables relativement comparables à celles de [6] et [7] pour classer les individus en fonction des caractéristiques de leur organisation du travail. Afin d'avoir quelques éléments comparatifs, nous avons repris le champ qu'il proposait dans leurs travaux, c'est à dire des individus salariés dans des entreprises de plus de 10 salariés dans tous les secteurs à l'exception de l'agriculture, les services à la personne les services non marchand public (santé, éducation, social...).

3.2. La typologie des modes d'organisation du travail

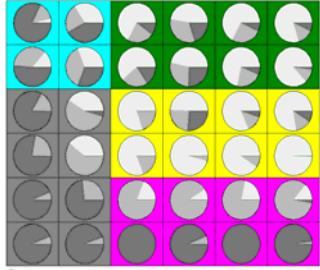
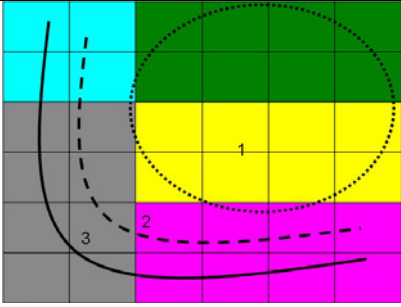
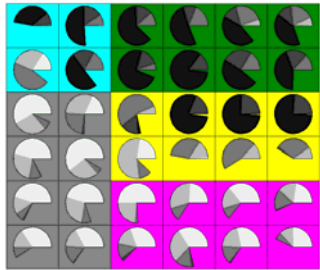
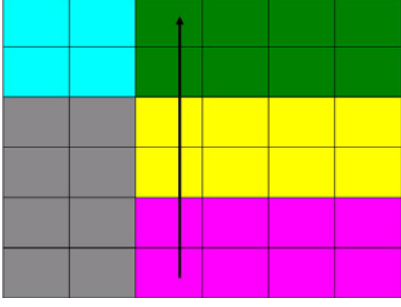
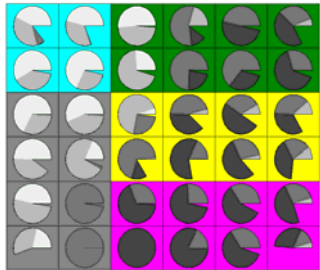
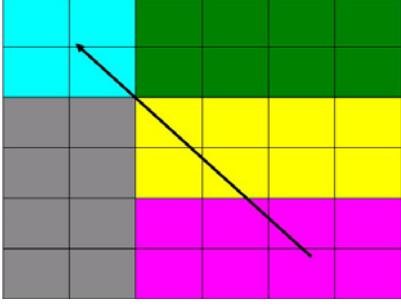
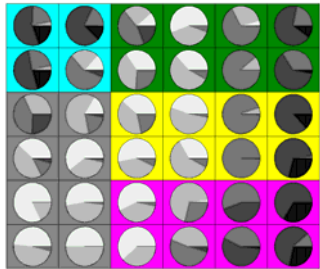
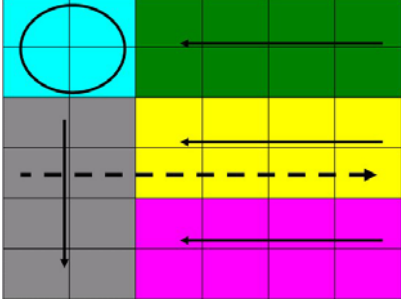
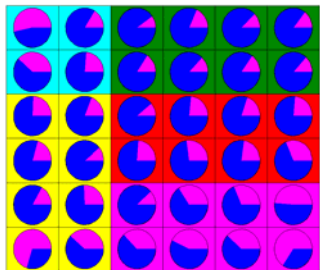
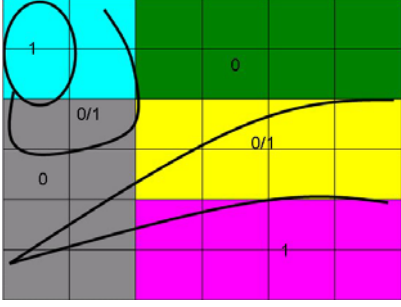
Nous introduisons plusieurs types de variables disponibles dans l'enquête « Génération 98 » et relatives aux modes d'organisation du travail. Ces modes d'organisation ne concernent pas forcément l'ensemble l'entreprise mais comme dans [6] et [7], l'équipe de travail dans laquelle sont insérés les individus. Les principales variables introduites sont : des variables sur le contrôle des individus par leur supérieur hiérarchique - existence et fréquence du contrôle du supérieur hiérarchique sur le travail effectué, intervention du supérieur hiérarchique pour répartir le travail entre les collègues de travail -, des variables

liées à l'autonomie de l'individu - la possibilité de négocier son planning de travail, l'existence d'objectif à moyen ou long terme -, une variable liée à la gestion de la qualité (la possibilité de proposer des méthodes ou des procédures pour améliorer son travail) -, des variables liées au contenu cognitif du travail - avoir à résoudre des problèmes imprévus, acquisition de nouvelles compétences dans l'exercice de leur activité - et une variable liée au travail en équipe. Dans la suite, nous nous concentrons sur la contribution à la typologie des variables quantitatives pour illustrer la problématique du papier. Cinq variables sur dix-huit ont un caractère quantitatif. Leur codage initial rend compte d'une approche quantitative, hormis un éventuel code pour non concerné qui n'est pas toujours facile à inscrire dans l'évolution quantitative. Il nous est alors arrivé de considérer la variable comme manquante dans le cas « non concerné » et d'ajouter une nouvelle variable codant le caractère « non concerné » (les variables étant par ailleurs toutes renseignées, cela ne génère pas d'ambiguïté). Par exemple, le travail en équipe est codé de 1 à 4 et manquant pour, respectivement, - plus de 50%, 50%, 25%, peu de travail et pas de travail en équipe (le saut de 1 vaut 25% et une variable supplémentaire vaut 0 ou 1 selon qu'il n'y a pas de travail en équipe)-, les objectifs de travail sont codés de 1 à 5 et manquant pour, respectivement, - à la journée, semaine, mois, trimestre, année et pas d'objectif fixé -.

Une classification des centres de classes par une classification hiérarchique avec la distance de Ward permet de distinguer 5 zones sur la carte. La projection des variables quantitatives permettent dans certains cas d'interpréter la position relative des ces zones par rapport aux évolutions de ces grandeurs et dans d'autres cas à interpréter la position des classes à l'intérieur de leur zone d'affectation. Sur les figures du tableau 4, le niveau de gris des unités renseigne sur le deuxième niveau de classe. Les graphiques du tableau 4 représentent le croisement des classes avec des variables endogènes comme expliqué en section 3. Concernant les variables liées à l'autonomie des individus, la carte (tableau 4, ligne 1) oppose les individus qui n'ont pas la possibilité de fixer leur planning de travail (partie nord-est de la carte) à ceux qui en ont toujours la possibilité (côtés et angles opposés). La carte oppose aussi les objectifs à courte et longue échéance (opposition sud-nord) (tableau 4, ligne 2). L'axe sud-est nord-ouest hiérarchise la part du travail en équipe (tableau 4, ligne 3). Concernant les rapports avec le supérieur hiérarchique, une opposition assez nette apparaît d'abord entre les rapports fréquents et les rapports beaucoup plus rares, expliquant à la fois les différences inter-zones et intra-zones (tableau 4, ligne 4). Ainsi la carte oppose des zones où le supérieur hiérarchique est plus présent en général (ouest) mais aussi les classes à l'intérieur d'une zone. L'accessibilité et le dirigisme du supérieur hiérarchique pouvant chacun créer une fréquence de rapports l'interprétation dépend du contexte, (définir le contexte d'une classe revient à définir sa zone). Les individus qui n'ont jamais à résoudre des problèmes imprévus occupent l'axe sud-ouest nord-est (tableau 4, ligne 5). L'apport de la méthode proposée, dans le cadre de cette application se traduit par une lecture très facile de la répartition sur la carte par l'évolution des grandeurs. La répartition des classes de bas en haut, de gauche à droite ou en diagonale sur la carte comme sur une zone s'explique très simplement par la hiérarchisation des grandeurs quantitatives.

Tableau 4

Représentations graphiques des résultats. Croisement des classes avec les modalités des variables qualitatives (ici endogènes).

<p>Trois couches successives d'individus qui n'ont jamais, parfois, toujours la possibilité de fixer leur planning de travail (codées respectivement 1, 2, 3).</p>	 <p>variable: x8n modalités: 1 → 3</p>	
<p>Oppose des objectifs à courte et longue échéance suivant l'axe sud-nord.</p>	 <p>variable: x9n modalités: 1 → 5</p>	
<p>L'axe sud-est nord-ouest hiérarchise la part du travail en équipe.</p>	 <p>variable: travequie modalités: 1 → 4</p>	
<p>La fréquence des rapports avec le hiérarchique explique à la fois les différences inter-zones et intra-zones.</p>	 <p>variable: x6n modalités: 1 → 5</p>	
<p>Les individus qui n'ont jamais à résoudre des problèmes imprévus occupent l'axe sud-ouest nord-est.</p>	 <p>variable: l_probim modalités: 0 → 1</p>	

4. CONCLUSION

Nous avons présenté une méthode très simple à mettre en œuvre pour prendre en compte simultanément les variables qualitatives et quantitatives. Cette transformation des dimensions quantitatives afin de les traiter comme des grandeurs qualitatives en gardant leurs propriétés est très complémentaire des cartes d'auto-organisation. Les exemples traités ont montré l'apport du codage aux cartes d'auto-organisation. La continuité d'une distribution et l'ordonnement de la grandeur sont en particulier bien rendus par le système de représentation associé. Elle correspond de plus à la nécessité lorsqu'on utilise un modèle non linéaire de conserver chez les dimensions traitées leur caractéristique de non linéarité. La simplicité de ce codage pourrait généraliser le traitement simultané des grandeurs qualitatives et quantitatives et ainsi, éviter l'usage regrettable de considérer qu'une enquête doit être établie à partir d'une approche quantitative ou qualitative.

BIBLIOGRAPHIE

- [1] M. Cottrell., S. Ibbou, P. Letrémy et P. Rousset (2006), Cartes auto-organisatrices de Kohonen, in Apprentissage connexionniste. Y. Bennani Ed., Hermès, Paris, p. 141-184.
- [2] J.C. Fort (2006). SOM's mathematics , Neural Networks, 19, 812-816.
- [3] T. Kohonen (2001). Self-Organizing Maps. 3.ed, Springer Series in Information Sciences, 30, Springer Verlag, Berlin.
- [4] L. Lebart , M. Morineau et M. Piron (2006). Statistique exploratoire multidimensionnelle. 4e édition, Dunod, Paris.
- [5] B. Escoffier-Cordier, Traitement simultané de variables qualitatives et quantitatives en analyse factorielle, Analyse des correspondances, Presses Universitaires de Rennes (2003) pp.65-74.
- [6] E. Lorenz, A. Valeyre (2005), « Les formes d'organisation du travail dans les pays de l'Union européenne », Travail et Emploi, n 102, avril-juin, pp. 91-105.
- [7] E. Lorenz, A. Valeyre (2005) « Organisational Innovation, Human Resource Management and Labour Market Structure : a Comparison of the EU-15», The Journal of Industrial Relations, Vol. 47, December, pp. 424-442.
- [8] E. Miret, F. García-Lagos, G. Joya, H. Arazoza & F. Sandoval (2005). A combined multidimensional scaling + self-organizing maps method for exploratory analysis of qualitative data. Proceedings of the 5th Workshop on Self-Organizing Maps (WSOM'05), Paris, France, 711-718.
- [9] J.B Mac Queen (1967) Some methods for classification and analysis of multivariate observations. Proceedings Symp. Math. Statist. And probability (5th), Berkley, 1, pp.281-297, University of Californie. Press, Berkley.

ISSN : 1776-3177
Marseille, 2008.