

Adaptation de l'algorithme SOM à l'analyse de données temporelles et spatiales : Application à l'étude de l'évolution des performances européennes en matière d'emploi

Catherine Aaron, Corinne Perraudin, and Joseph Rynkiewicz

SAMOS - MATISSE

Université de Paris 1

90 rue de Tolbiac,

75013 Paris, France

(e-mail: catherine.aaron@hotmail.com, Corinne.Perraudin@univ-paris1.fr,

Joseph.Rynkiewicz@univ-paris1.fr)

Abstract. Cet article étudie l'évolution des performances européennes en matière d'emploi depuis le début des années quatre vingt dix, en utilisant l'algorithme SOM adapté au traitement de données qui sont à la fois temporelles et spatiales. La carte de Kohonen ainsi obtenue permet d'établir une classification des pays de l'Union Européenne qui tient compte simultanément de l'ordonnement temporel et spatial des données, et permet alors de comparer les trajectoires des différents pays dans le temps. Nous comparons les résultats obtenus par cette méthode à ceux reposant sur une carte de Kohonen traditionnelle.

Keywords: Classification, Algorithme SOM, Emploi, Union Européenne.

1 Introduction

Malgré les modestes performances européennes en matière d'emploi très souvent dénoncées, la question de l'emploi n'a véritablement été abordée au niveau européen qu'en 1993 avec le fameux "Livre blanc Delors" sur la croissance, la compétitivité et l'emploi. Inspirés de ce document, les Conseils européens qui suivent vont intégrer progressivement l'objectif d'atteindre un niveau élevé d'emploi parmi les objectifs clés de l'Union Européenne (UE ensuite), conduisant à lui donner le même niveau d'importance que les objectifs macroéconomiques de croissance et de stabilité. Lors du sommet sur l'emploi de 1997 à Luxembourg, la Stratégie Européenne pour l'Emploi (SEE) est lancée. Elle est conçue comme un instrument de coordination et d'orientation des priorités de la politique de l'emploi des Etats membres de l'UE. Le Conseil européen de Lisbonne, en mars 2000, lui donne une impulsion supplémentaire en insérant la stratégie dans l'agenda global de l'UE en matière économique et sociale. Il vise à faire de l'UE l'économie "la plus compétitive et la plus dynamique du monde" et établit des cibles à atteindre à un horizon de dix ans.

A mi-chemin de la date butoir, il importe d'étudier la situation en matière d'emploi des pays de l'UE, et de regarder plus particulièrement si le fait d'avoir mis l'emploi au cœur des préoccupations européennes a permis une amélioration des performances en matière d'emploi. Cet article étudie l'évolution des performances des pays de l'UE en matière d'emploi depuis le début des années quatre vingt dix. Il s'agit de tenir compte de l'évolution temporelle de différents indicateurs relatifs au marché du travail afin d'étudier les trajectoires suivies par les pays de l'UE.

Nous étudions la situation des 25 pays actuellement dans l'UE, ainsi que celles de la Roumanie et de la Bulgarie, qui vont rejoindre l'Union Européenne en 2007, de la Turquie, dont les négociations devraient débuter fin 2005 et de la Croatie, candidate à l'adhésion. Nous retenons, dans une première analyse, les trois indicateurs définissant les objectifs d'emploi fixés par la Stratégie de Lisbonne à atteindre pour 2010, à savoir un taux d'emploi total de 70%, un taux d'emploi des femmes de 60% et un taux d'emploi des travailleurs âgés (55-64 ans) de 50%. Nous étendons ensuite cette analyse à un ensemble plus large d'indicateurs relatifs à l'emploi et au chômage, afin de caractériser de manière plus fine les performances des marchés du travail européens.

Pour mener à bien cette comparaison des trajectoires des pays européens selon leurs performances, nous proposons d'utiliser une adaptation de l'algorithme SOM, permettant de traiter des données qui sont la fois temporelles et spatiales. La section suivante présente l'adaptation de l'algorithme SOM. La troisième section présente les résultats de l'analyse des taux d'emploi. La quatrième section étend cette analyse à une gamme plus large d'indicateurs du marché du travail. La dernière section conclut.

2 Carte de Kohonen adaptée à l'analyse de données temporelles

L'algorithme SOM (*Self-Organizing Map*), encore appelé algorithme de Kohonen, est un algorithme stochastique de classement des données, qui respecte la topologie de l'espace des observations en intégrant une notion de voisinage entre les classes (voir [Kohonen, 1995])¹.

Afin de classifier les données européennes étudiées dans cet article, qui sont temporelles ($t = 1, \dots, T$) et spatiales ($j = 1, \dots, N$), on peut considérer qu'un pays j pour une année t est une observation et on peut alors classer les $T \times N$ observations sur une carte de Kohonen². Cependant, cette manière de faire ne permet pas d'observer les trajectoires des individus et de les comparer facilement.

¹ Une présentation des méthodes de classification et de diverses applications construites à partir de cet algorithme est fournie par [Cottrell *et al.*, 2003].

² Voir par exemple [Akarçay-Gürbüz and Perraudin, 2004] pour une classification des pays européens selon leurs performances économiques dans le temps

Afin de réellement prendre en compte la dimension temporelle des données, on pourrait alors construire T cartes de Kohonen et classifier les N observations selon les variables retenues par année. Le problème avec ce type de méthode est que la classification ainsi obtenue année par année est très instable.

Ainsi, nous proposons une méthode qui permet de tenir compte de l'aspect temporel des données³. Plus précisément, nous adaptons l'algorithme SOM afin qu'il tienne compte de l'ordonnement temporel en plus de l'ordonnement spatial des données. Pour cela, nous construisons une carte dont la longueur est égale au nombre d'années T de la période d'observations, et de largeur égale au nombre de représentants par année choisi a priori (noté k). Le calcul des vecteurs codes de cette carte de Kohonen, dite généralisée dans la suite, s'effectue selon l'algorithme suivant :

- L'initialisation de l'algorithme SOM correspond à un tirage aléatoire de k pays dans l'ensemble des données. A l'unité (i, t) de la carte, on affecte les valeurs des variables⁴ du pays i pour l'année t .
- A chaque itération, un pays i_0 et une année t_0 sont tirés aléatoirement dans l'ensemble de données. Ensuite, pour tout $i \in [1, k]$, on cherche l'unité (i, t_0) qui est la plus proche de l'observation sélectionnée.
- On met à jour l'unité gagnante et les unités voisines. Le voisinage décroît dans la dimension ligne durant les itérations de r à 0. Pour forcer l'organisation temporelle, pour un voisinage ligne donné r , la taille du voisinage temporel décroît de r à 0 (voir figure 1).
- Finalement, afin de garantir la convergence, on finit à 0 voisin sur les deux derniers tiers des itérations.

Une fois que l'algorithme a convergé, on positionne les pays sur la carte afin d'identifier leur position.

Un premier avantage de cette carte de Kohonen généralisée au classement de données temporelles est qu'elle permet d'observer une continuité à la fois dans la dimension temporelle et dans la dimension ligne.

Le nombre de colonnes de cette carte étant choisi a priori (arbitrairement grand), nous mettons ensuite en œuvre une classification ascendante hiérarchique à l'ensemble des vecteurs codes de la carte afin de réduire le nombre de classes. Ainsi, le nombre de super-classes par année n'est pas contraint à être le même au cours du temps. Cela nous permet d'étudier si les super-classes homogènes se forment, dans le temps ou alors pour une date donnée, et si le nombre de super-classes par année se réduit, indiquant alors une convergence des pays.

³ Voir [Aaron *et al.*, 2003], qui proposent deux adaptations de l'algorithme de Kohonen afin de classifier les trajectoires des 15 pays de l'UE dans le temps, vers les normes fixées par les critères de Maastricht. La première méthode proposée est reprise dans cet article.

⁴ Les données sont centrées et réduites par variable sur la période entière.

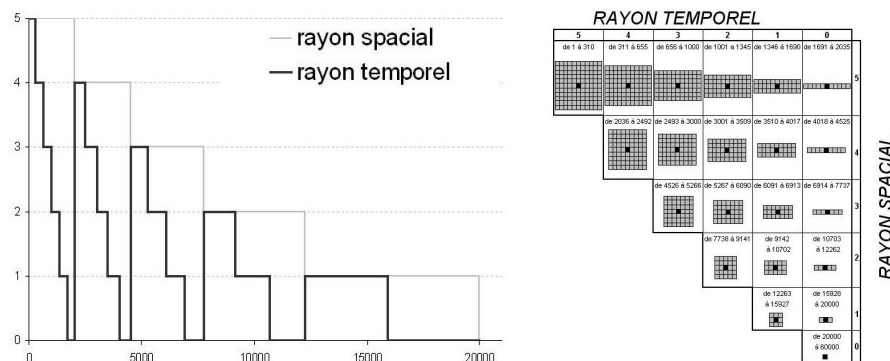


Fig. 1. Evolution du rayon dans l'algorithme généralisé.

Un deuxième avantage de cette carte est de permettre de visualiser très facilement les trajectoires des pays à travers les différentes super-classes.

3 L'évolution des taux d'emploi

Nous étudions tout d'abord les performances des pays européens sur la base du taux d'emploi total (TxEmp), du taux d'emploi des femmes (TxEmpF) et du taux d'emploi des travailleurs âgés (TxEmpTA)⁵.

Le taux d'emploi total moyen sur tous les pays disponibles et sur la période 1992-2003 s'élève à 62% (respectivement 53% pour les femmes et 37,5% pour les travailleurs âgés). Il passe de 62,4% en 1992 (respectivement 52% et 39%) à 62% en 2003 (respectivement 54,3% et 39,9%)⁶. Cependant, cette apparente constance des taux d'emploi cache des disparités importantes entre les pays, que ce soit en termes de niveaux ou de trajectoires, comme nous allons le voir.

Nous avons fixé à 12 le nombre de classes par année. La carte de Kohonen obtenue (voir figure 2) comprend alors 12 lignes (correspondant aux 12

⁵ Le taux d'emploi est égal au nombre de personnes en emploi rapporté à la population concernée en âge de travailler. Les données sont issues de [Commission, 2004] et du site <http://europa.eu.int>. Les variables sont disponibles de 1992 à 2003 pour l'Allemagne (DE), la Belgique (BE), le Danemark (DK), l'Espagne (ES), la Finlande (FI), la France (FR), la Grèce (EL), l'Irlande (IE), le Luxembourg (LU), les Pays-Bas (NL), le Portugal (PT), le Royaume-Uni (UK), la Suède (SE). Les données ne sont disponibles qu'à partir de 1993 pour l'Italie (IT); de 1994 pour l'Autriche (AT); de 1996 pour la Hongrie (HU), la Slovénie (SI); de 1997 pour la Pologne (PL), la Roumanie (RO); de 1998 pour la République tchèque (CZ), l'Estonie (EE), la Lettonie (LV), la Lituanie (LT), la Slovaquie (SK); de 2000 pour la Bulgarie (BG), Chypre (CY), Malte (MT), la Turquie (TR); et seulement pour 2003 pour la Croatie (CR).

⁶ La moyenne étant effectuée sur 13 pays en 1992 et sur 29 en 2003.

années de la période 1992-2003) et 12 colonnes. On observe une continuité dans la répartition des trois indicateurs à la fois dans la dimension ligne et dans la dimension colonne. Chaque année, les pays sont classés selon leurs performances en matière de taux d'emploi. On constate que la carte oppose les pays ayant les meilleures performances en termes de taux d'emploi, que ce soit total, des femmes ou des travailleurs âgés (le coté gauche de la carte) et les pays ayant de moins bonnes performances (coté droit de la carte). On remarque que le coin supérieur droit de la carte correspond aux taux d'emploi total et féminin les plus faibles, mais à des taux d'emploi moyens pour les travailleurs âgés. On note que les différences de performances se sont estompées dans le temps, essentiellement parce que les moins bonnes performances en matière de taux d'emploi se sont améliorées.

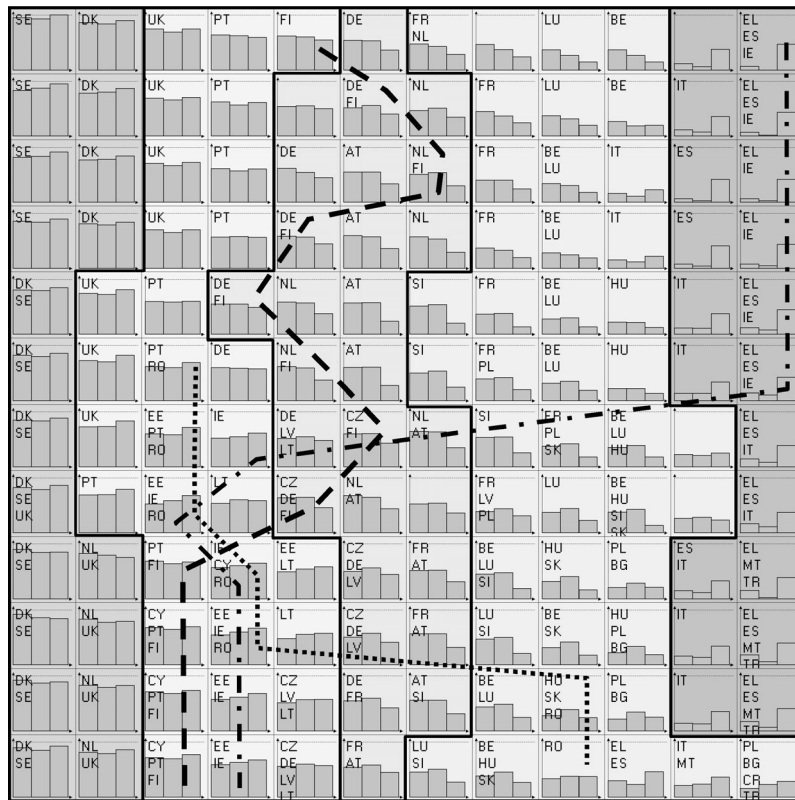


Fig. 2. Carte de Kohonen généralisée.

Note : La première ligne correspond à 1992 et la dernière à 2003. Dans chaque case de la carte sont représentées de gauche à droite les répartitions de TxEmp, TxEmpF, TxEmpTA. La ligne en pointillés (respectivement discontinue et alternée) représente la trajectoire de la Roumanie (respectivement de la Finlande et de l'Irlande).

La classification ascendante hiérarchique menée sur l'ensemble des vecteurs codes de la carte conduit à retenir 5 super-classes. Elles correspondent à des performances de moins en moins bonnes en allant de la gauche vers la droite. On remarque que les 5 super-classes coexistent sur l'ensemble de la période étudiée (sauf celle regroupant les pays aux performances médiocres qui disparaît la dernière année). La super-classe qui se trouve sur le côté gauche de la carte est la seule à enregistrer des taux d'emploi supérieurs aux cibles fixées pour 2010. Il s'agit des pays nordiques (Suède, Danemark, ainsi que Pays-Bas depuis 2000) ainsi que le Royaume-Uni. La super-classe à droite de la carte correspond aux moins bonnes performances, et on y retrouve les pays du Sud (Espagne, Grèce, Italie, ainsi que Malte, Turquie, Bulgarie, Croatie). La super-classe du milieu correspond à des performances moyennes et regroupe la France, l'Allemagne, l'Autriche, ainsi que les autres pays récemment entrés dans l'UE.

On peut très facilement observer la trajectoire suivie par chaque pays à travers la carte, et donc à travers les super-classes décrivant des performances différentes, et cela grâce à l'ordonnement temporel imposé dans cette carte. Certains pays restent tout au long de la période dans la même super-classe. C'est le cas de la Suède et du Danemark, qui enregistrent les meilleures performances, ou de l'Espagne, l'Italie et la Grèce qui restent dans la super-classes des moins bonnes performances. Quelques pays suivent des trajectoires qui traversent différentes super-classes, indiquant une évolution, soit vers de meilleures performances (Irlande ou Finlande, dont les trajectoires sont dessinées sur la carte) soit vers des performances en déclin (cas de la Roumanie, dont la trajectoire est aussi représentée sur la carte).

Afin de souligner l'avantage de cette carte relativement à une carte de Kohonen non contrainte par l'ordonnement temporel (dite traditionnelle dans la suite)⁷, nous comparons ces résultats avec ceux obtenus en considérant un pays pour une année comme une observation.

Dans ce cas, nous obtenons une classification (voir carte figure 3) qui associe essentiellement un même pays pour différentes années. Comme dans les résultats précédents, cette carte permet de séparer très clairement les pays qui enregistrent de bonnes performances en matière de taux d'emploi (en haut à droite de la carte figure 4) aux pays ayant les moins bonnes performances (à gauche de la carte). Parmi ces derniers, on retrouve la distinction entre ceux qui ont des taux d'emploi des travailleurs âgés les plus faibles et ceux qui ont des taux moyens.

Le positionnement des pays, ainsi que les super-classes obtenues par une classification hiérarchique ascendante (en 5 super-classes), gardent le même type d'interprétation que dans les résultats précédents, mais il est moins aisé d'observer le suivi des trajectoires et de les comparer entre elles (voir carte figure 3). A titre d'illustration, nous avons reproduit sur la carte 3 la trajectoire des trois pays (Roumanie, Finlande, Irlande) mentionnés précédemment.

⁷ Voir [Letrémy, 2000] pour une présentation des programmes en SAS-IML.

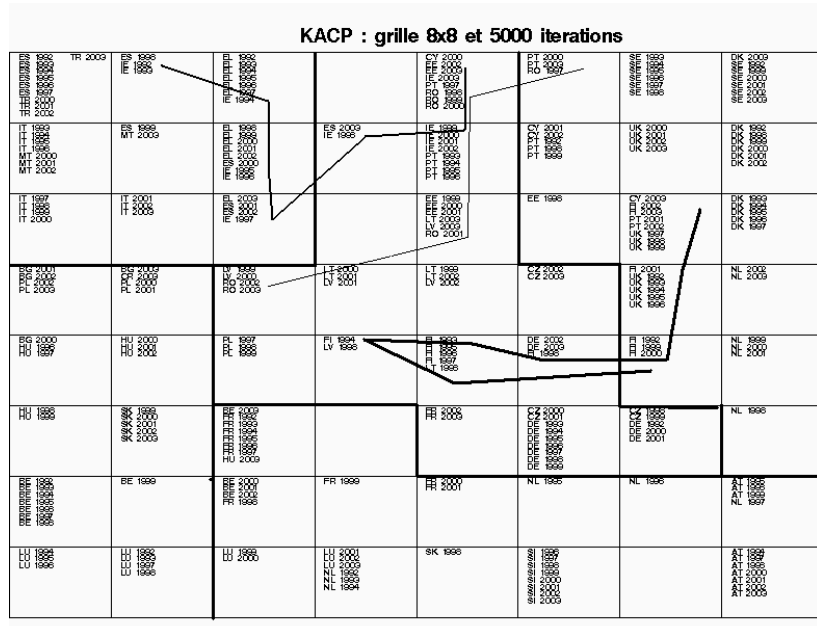


Fig. 3. Carte de Kohonen traditionnelle : répartition des pays.

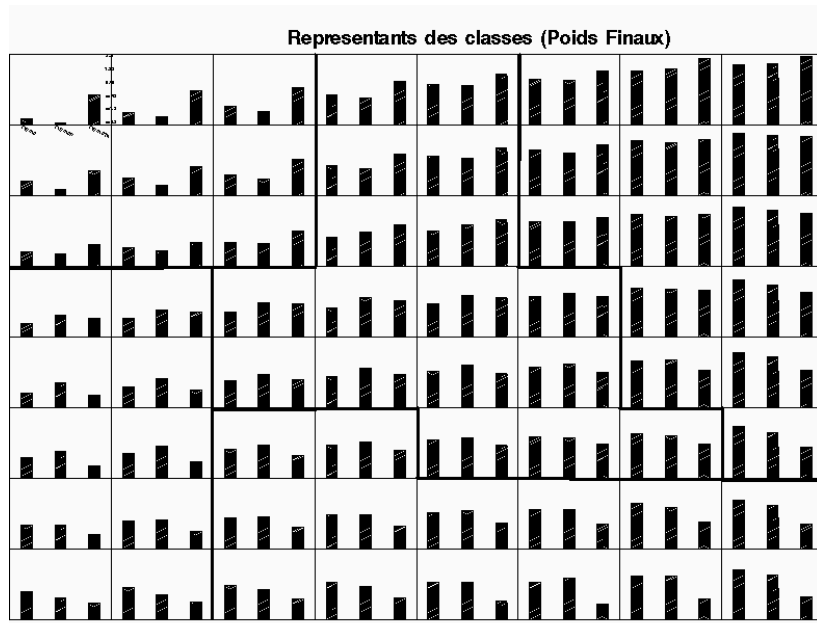


Fig. 4. Carte de Kohonen traditionnelle : répartition des variables.

4 L'évolution des taux d'emploi et d'autres indicateurs

Nous étendons dans cette section l'analyse menée sur les taux d'emploi à un ensemble plus large d'indicateurs relatifs à l'emploi et au chômage, afin de caractériser de manière plus fine les performances des marchés du travail européens. Nous considérons en plus des taux d'emploi total, des femmes et des travailleurs âgés, le taux d'emploi des jeunes (15-24 ans), le taux de croissance de l'emploi, le taux de chômage, le taux de chômage de longue durée (plus de 12 mois) et le taux de contrats à durée déterminée (CDD). La carte obtenue (figure 5) indique que la méthode proposée dans cet article est robuste à un ensemble plus important de variables.



Fig. 5. Carte de Kohonen généralisée.

Note : La première ligne correspond à 1992 et la dernière à 2003. Dans chaque case de la carte sont représentées de gauche à droite les répartitions de TxEmp, TxEmpF, TxEmpTA, TxEmpJ, CroissEmp, Cho, ChoLD, CDD.

La classification hiérarchique ascendante conduit à retenir 7 super-classes. La super-classe se trouvant à gauche de la carte correspond encore aux pays enregistrant les meilleures performances en matière de taux d'emploi, de croissance de l'emploi et les plus faibles taux de chômage, combinés avec des taux moyens de CDD. On retrouve les pays nordiques (Suède, Danemark, Pays-Bas ainsi que Royaume-Uni). Les deux super-classes à coté correspondent à des performances moyennes : on trouve l'Allemagne, l'Autriche, la Finlande à la fin de la période. La super-classe en haut à droite de la carte, correspondant à des taux d'emploi très faibles et des taux de chômage très élevés (Espagne), disparaît en 1998. La classe en bas à droite est principalement caractérisée par des taux de chômage total et de longue durée très élevés, on y trouve des pays entrés récemment dans l'UE (Pologne, Slovaquie, Lituanie, ainsi que la Bulgarie). La super-classe aux performances médiocres où se trouve l'Espagne comprend aussi la Turquie dans les années 2000. On note les meilleures performances de l'Estonie, de la République tchèque ou même de la Hongrie.

5 Conclusion

A travers l'étude des performances en matière d'emploi des pays européens et des pays candidats à l'adhésion, cet article a illustré les avantages d'une carte de Kohonen adaptée afin de prendre en compte la dimension temporelle de données. Cette méthode permet notamment de suivre très facilement les trajectoires suivies par les pays.

References

- [Aaron *et al.*, 2003]C. Aaron, C. Perraudin, and J. Rynkiewicz. Curves based kohonen map and adaptative classification : an application to the convergence of the european countries. Prepub SAMOS 190, 2003.
- [Akarçay-Gürbüz and Perraudin, 2004]A. Akarçay-Gürbüz and C. Perraudin. How to situate the turkish economy among the european union economies? an exploratory analysis. *European Journal of Economics and Social Systems*, pages 41–62, 2004.
- [Commission, 2004]European Commission. *Employment in Europe 2004*. European Communities, 2004.
- [Cottrell *et al.*, 2003]M. Cottrell, S. Ibbou, P. Letrémy, and P. Rousset. Cartes auto-organisées pour l'analyse exploratoire des données et la visualisation. *Journal de la société française de statistique*, pages 67–106, 2003.
- [Kohonen, 1995]T. Kohonen. *Self-Organizing Maps*. Springer, 1995.
- [Letrémy, 2000]P. Letrémy. Notice d'installation et d'utilisation de programmes basés sur l'algorithme de kohonen et dédiés à l'analyse des données. Prepub SAMOS 121, 2000.