

# Algorithme de Kohonen : application à l'analyse de données économiques

Autor(en): **Blayo, François / Demartines, Pierre**

Objektyp: **Article**

Zeitschrift: **Bulletin des Schweizerischen Elektrotechnischen Vereins, des Verbandes Schweizerischer Elektrizitätsunternehmen = Bulletin de l'Association Suisse des Electriciens, de l'Association des Entreprises électriques suisses**

Band (Jahr): **83 (1992)**

Heft 5

PDF erstellt am: **11.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-902800>

## **Nutzungsbedingungen**

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

## **Haftungsausschluss**

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

# Algorithme de Kohonen

## Application à l'analyse de données économiques

François Blayo, Pierre Demartines

**L'algorithme d'auto-organisation de Kohonen est un algorithme neuronal qui permet de réduire la dimension d'un problème sans connaissance a priori sur ses caractéristiques. Il est appliqué ici à un problème de traitement de données économiques et comparé à une méthode de projection orthogonale traditionnelle: l'analyse en composantes principales.**

**Der neuronale Selbstorganisations-Algorithmus von Kohonen erlaubt, die Dimension eines Problems ohne A-priori-Kennntnis seiner Charakteristiken zu reduzieren. In diesem Beitrag wird die neuronale Methode auf die Verarbeitung von ökonomischen Daten angewendet und mit der traditionellen Orthogonalprojektions-Methode, der Analyse mit Hilfe von Eigenvektoren, verglichen.**

La simple observation d'un phénomène ne peut suffire à le comprendre profondément (c'est-à-dire à dégager des relations structure, causalité, etc.) qui existent entre tous les éléments qui constituent l'image du phénomène. Les observations sont seulement une liste de données qui sont précisément cette image. Son analyse et la synthèse des déductions faites permettent de construire une compréhension partielle du phénomène, transformant ainsi les données en informations. Cette transformation ne peut pas toujours être aisément construite sur des données multiples et complexes. Elle nécessite une réduction de la complexité par des techniques appropriées (régression linéaire, analyse canonique, analyse discriminante) qui relèvent du domaine de l'analyse de données. Par exemple, les mesures économiques propres aux pays, mettent en jeu un grand nombre de données et requièrent un savoir spécialisé pour en déduire des informations pertinentes susceptibles d'être exploitées par des décideurs (chefs d'entreprises, hommes politiques, etc.).

Les réseaux de neurones artificiels (RNA) montrent une certaine capacité pour l'analyse de données. Ils ont déjà été appliqués à des problèmes de prédiction de séries temporelles [1], de classification [2], d'analyse de régressions non-paramétriques [3]. Actuellement, les techniques employées dans le domaine de l'analyse de données sont bien établies et sont éprouvées sur de nombreux exemples concrets. Il est donc intéressant, à titre de comparaison, d'appliquer une technique d'analyse à base de RNA ainsi qu'une technique classique d'analyse de données à un même problème afin de comparer les résultats obtenus par ces différentes méthodes.

Dans les lignes qui suivent, l'Analyse en Composantes Principales (ACP) est ainsi comparée à l'algorithme d'auto-organisation de Kohonen. Une telle démarche a déjà été utilisée pour comparer l'ACP à l'Analyse en Composantes Indépendantes [4]. Mais une comparaison ne doit pas s'arrêter à la simple fonction réalisée par l'algorithme. Elle doit également prendre en compte la vitesse et la parallélisation possible du calcul, et la possibilité d'adapter le traitement à de nouvelles données.

### L'Analyse en Composantes Principales

L'ACP, une méthode classique d'analyse de données, est utilisée quand le nombre de variables (ou de caractères selon la terminologie employée en statistiques) est plus grand ou égal à 3, car une construction graphique directe est délicate à analyser ou même impossible à représenter. Sans décrire le détail de la méthode, rappelons que l'ACP effectuée une projection orthogonale d'une distribution N-dimensionnelle ( $N > 3$ ) sur le plan principal défini par des vecteurs propres de la matrice de covariance de la distribution. Le calcul des vecteurs propres requiert une diagonalisation de matrice qui peut être très coûteuse en temps calcul.

L'ACP est donc une méthode de réduction de dimension d'un problème basée sur une composition des caractères entre eux et non sur une simple sélection, mais suivant des critères judicieusement choisis. Après transformation, la qualité de la projection effectuée par l'ACP peut être estimée pour dégager la quantité d'information conservée. A titre d'illustration, une distribution tridimensionnelle en forme de «fer à cheval» est

#### Adresse des auteurs

Dr. François Blayo et Pierre Demartines, ing. dipl. EPFL, Ecole Polytechnique Fédérale de Lausanne, Laboratoire de Microinformatique, INF - Ecublens, 1015 Lausanne.

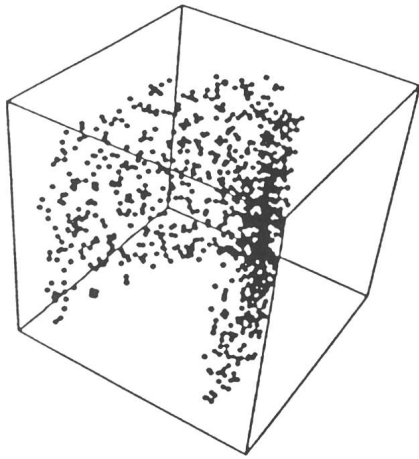


Figure 1 Exemple de distribution en «fer à cheval»

montrée sur la figure 1, et un plan principal calculé par l'ACP ainsi que la projection orthogonale (bi-dimensionnelle) de la distribution sur ce plan sont montrés sur les figures 2a et 2b.

### L'algorithme de Kohonen

L'algorithme de Kohonen [5] effectue également une projection, mais sur un sous-espace engendré par une grille discrète de neurones formels, situés à l'emplacement de leurs poids. Cette structure est appelée carte d'auto-organisation. La carte, pour laquelle les liens entre neurones sont fixés, va permettre d'établir une relation entre les points (ou individus) de la distribution servant à l'apprentissage et l'activité des neurones. Cette relation est établie par une transformation fortement non-linéaire, qui tend à respecter les relations de voisinage de la distribution ainsi que la densité ponctuelle de cette fonction de distribution. Ainsi, après un calcul itératif durant lequel on applique la règle d'adaptation, chaque neurone devient sensible à une portion de l'espace qui a servi à l'apprentissage. Une description détaillée de l'algorithme et illustrée de nombreux exemples peut être lue dans [5].

Pour illustration, on montre sur la figure 3b la forme prise par la grille du réseau à l'intérieur de la distribution en «fer à cheval», ainsi que la projection de cette distribution sur la surface gauche formée par la grille de la figure 3a. On voit clairement que la transformation est une projection non-linéaire. La grille bi-dimensionnelle se déforme pour construire une surface gauche telle que chaque neurone se

spécialise pour représenter une portion de la distribution.

### Simulations et résultats

On constate d'emblée que la répartition des données projetées sur le plan principal avec l'ACP produit une répartition des points non uniforme, comme cela est montré sur la figure 2b. La carte de Kohonen, comme le montre la figure 3a, va s'inscrire dans la distribution, et calcule une projection de la distribution. Chaque neurone, représenté par un point à l'intersection de la grille, est ainsi sensible à une portion de l'espace d'entrée dans laquelle la répartition des points tend vers une distribution uniforme. De plus, s'il existe une relation de voisinage entre les individus de la distribution (au sens de n'importe quelle mesure de distance), alors cette relation est conservée dans la grille de la carte

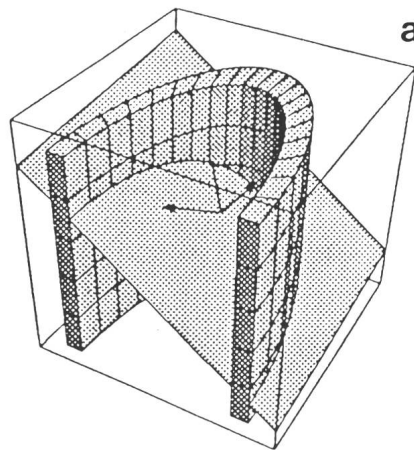


Figure 2a Un plan principal obtenu par l'ACP

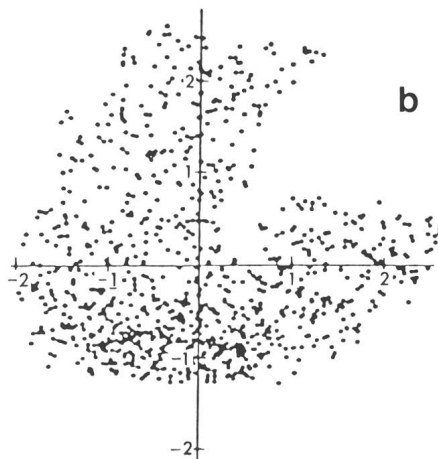


Figure 2b Distribution projetée sur le plan principal

de Kohonen. Ainsi, après apprentissage, la présentation au réseau de deux points proches (en distance Euclidienne par exemple) dans la distribution, produira une activité de deux neurones proches sur la carte de Kohonen. Cette notion d'ordonnement est essentielle dans le fonctionnement de l'algorithme, mais sa définition précise n'est pas encore arrêtée. La figure 3b montre la projection obtenue avec un réseau de Kohonen.

L'ACP, comme cela a été dit précédemment, est une méthode de projection linéaire (orthogonale), alors que l'algorithme de Kohonen est une méthode de projection non-linéaire qui tente d'inscrire un support discret représentant une surface gauche dans la distribution des données d'apprentissage. D'autre part, l'ACP et l'algorithme de Kohonen ont un point commun: la détection des variances élevées. Dans l'exemple d'une distribution en fer à cheval, les variances selon les différents axes de la distribution sont comparables. Il y a donc, avec l'ACP, une infinité de solutions, chacune correspondant à un plan principal différent. Avec l'algorithme de Kohonen, il y aura également une détection des variances élevées, mais selon des coordonnées définies par la topologie du réseau.

La méthode de calcul est un facteur de comparaison qui doit être pris en compte. L'ACP est une méthode globale, qui implique le calcul d'une inversion de matrice. De plus, tout ajout d'un ou plusieurs éléments nouveaux engendre un nouveau calcul de cet inverse. L'algorithme de Kohonen est une méthode itérative, qui ne requiert pas d'inversion de matrice. L'ajout de nouveaux éléments peut être pris en compte à tout moment, pour produire la carte. De plus le calcul peut être naturellement distribué sur une machine parallèle, ce qui n'est pas facilement réalisable avec l'ACP. L'algorithme de Kohonen est donc un bon candidat pour les futures machines basées sur des architectures de processeurs travaillant en parallèle.

Ces quelques éléments montrent que la comparaison entre algorithmes dits classiques et algorithmes neuro-naux ne peut être seulement faite sur des seules considérations de performances. Il faut également intégrer des facteurs tels que, par exemple, l'implantabilité sur machines parallèles, la possibilité de s'adapter à de nouvelles données, et surtout la différence fonctionnelle entre algorithmes.

## Application

A titre d'illustration, on applique les deux méthodes présentées à l'analyse de données économiques. Elles représentent l'état de 52 pays en 1984 par six caractères: la croissance annuelle, la mortalité infantile, le taux d'illettrisme, la fréquentation scolaire, le produit intérieur brut (PIB) par habitant (qui concerne les pays à économie planifiée) et la croissance annuelle du PIB. La base d'exemples est ainsi constituée de 52 prototypes, chacun d'eux étant mis sous la forme d'un vecteur de six valeurs, comme cela est montré dans l'exemple suivant:

France (0.4, 9.1, 1.2, 86.0, 11326.0, 5.1)

Après normalisation des données (moyenne nulle et variance égale à un) et calcul de la matrice de covariance, l'ACP peut être appliquée en calculant les valeurs propres et vecteurs propres de cette matrice. On choisit ensuite les axes principaux parmi ces vecteurs propres, et la projection orthogonale des vecteurs de dimension 6 sur le plan principal est calculée. Ceci donne le résultat montré sur la figure 4, pour laquelle la base du plan principal est constituée des deux vecteurs propres associés aux deux valeurs propres les plus grandes.

L'ACP produit un résultat très significatif: les projections des pays font apparaître des groupes correspondant à différentes situations géopolitiques. On distingue nettement les pays du tiers monde, les pays producteurs de

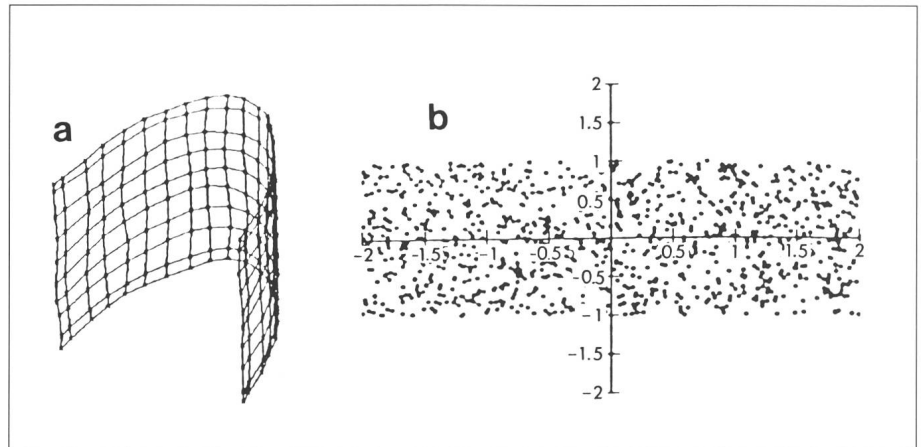


Figure 3a Grille obtenue avec l'algorithme de Kohonen

Figure 3b Projection obtenue par l'algorithme de Kohonen

pétrole, les pays de l'Est et d'Amérique du Sud, l'Europe des douze et le groupe des sept pays les plus industrialisés.

Le même jeu de données est fourni comme base d'apprentissage au réseau de Kohonen, constitué de six entrées et de cent neurones. Chaque individu de la base est choisi aléatoirement (on suppose que chaque individu a la même probabilité d'être tiré) et on le présente au réseau. L'algorithme de calcul des poids est alors appliqué et l'opération est répétée un grand nombre de fois. Après apprentissage, on fixe les poids et on identifie l'emplacement de chaque individu sur la carte. Cela est réalisé en présentant séquentiellement chacun des 52 individus et en relevant le neurone dont l'activité est la plus élevée pour l'individu

concerné. Le résultat atteint après 35000 tirages pour l'apprentissage est montré sur la figure 5. Cette opération ne prend que deux minutes de calcul environ sur une station Sun 4.

On constate que l'algorithme produit une carte qui reflète les mêmes groupements, mais sur laquelle les pays sont uniformément distribués sur des emplacements fixes (100 emplacements correspondant à 100 neurones). Le traitement ainsi effectué par l'algorithme de Kohonen permet de mettre en évidence des informations contenues dans les données, mais selon une représentation qui diffère de celle obtenue avec l'ACP. Le contenu informatif dégagé par ces deux méthodes est ici comparable, mais les algorithmes appliqués pour l'obtenir sont très différents.

## Conclusion

L'algorithme de Kohonen est une méthode de calcul qui présente un grand intérêt dans le domaine de l'analyse de données. Il effectue une opération de quantification vectorielle non-supervisée, ainsi qu'une réduction de dimension. Ces opérations sont le résultat d'une projection non-linéaire qui a deux propriétés: l'adaptation à la fonction de densité de distribution des individus, ainsi que le maintien d'une relation topologique entre les individus et les neurones. Cet algorithme n'est cependant pas complètement compris. Il reste encore à déterminer par quel principe il forme la surface gauche représentée par la grille de neurones, et surtout quelle est la signification d'une grille ordonnée. Il est probable qu'à l'issue de ces travaux de recherche, cet algorithme

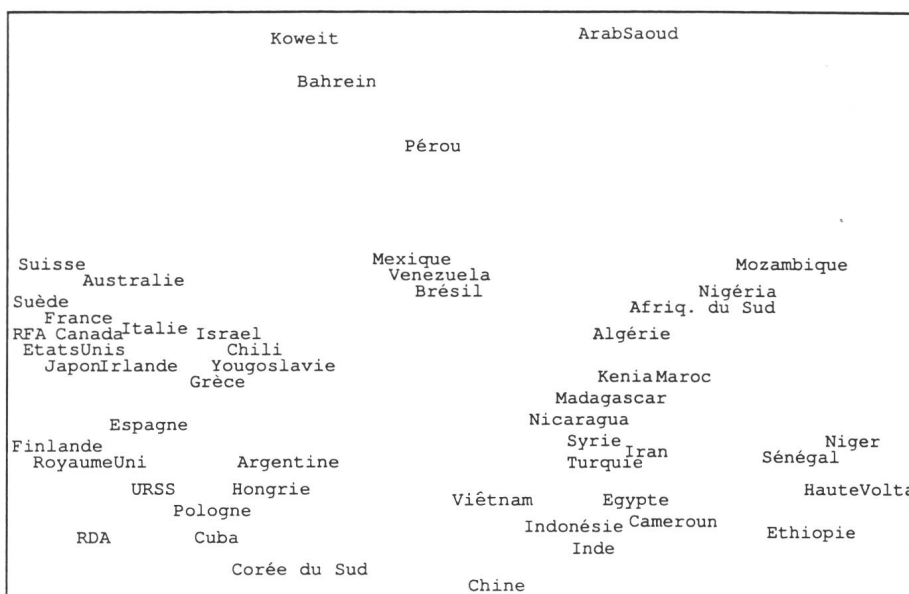


Figure 4 ACP sur les données économiques

pourra être correctement exploité car on en connaîtra précisément les avantages et les limites. En l'état, il consti-

tue une avancée importante dans la compréhension des phénomènes généraux d'auto-organisation.

**Remerciements**

Ce travail a été effectué dans le cadre d'un financement Fonds National pour le projet Esprit-BRA Nerves n° 3049.

Suède	France	Australie	Italie	Yougoslavie				
Suisse				Grèce		Koweït		
RFA	EtatsUnis							
	Japon		Irlande			Bahreïn		
	Canada							
Finlande		Espagne		Chili	Mexique	Pérou	ArabSa.	
URSS	RoyaumeUni		Israël	Venezuela	Brésil			
RDA								
Cuba		Argentine				Madagascar	Afriq. Mozamb.	
							du Sud Nigéria	
Corée du Sud	Pologne	Hongrie		Turquie	Maroc			
					Algérie			
Chine			Indonésie				Sénégal	
Viêtnam	Nicaragua	Syrie	Egypte	Cameroun				
		Kenia	Iran	Inde	Ethiopie	Niger	H.Volta	

Figure 5 Algorithme de Kohonen appliqué aux données économiques

**Bibliographie**

- [1] C. Windsor, A. Harker: Multi-variate financial index prediction – A neural network study. International Neural Networks Conference. 1. pp. 357–360. Paris, 1990.
- [2] T. Baumann, A. Germond, D. Tschudi: Impulse test fault diagnosis on power transformers using Kohonen's self organizing neural network. Proceedings of the Third Symposium on Expert System Applications to Power Systems. Tokyo & Kobe. April 1991.
- [3] V. Cherkassky, H. Lari-Najafi: Self-organizing neural networks for non-parametric regression analysis. International Neural Networks Conference. 1. pp. 370–374. Paris, 1990.
- [4] C. Jutten, J. Hérault: Independent Component Analysis (INCA) versus Principal Component Analysis. Signal Processing, 4. Theories and Applications. (1988) pp. 643–646. New-York. Elsevier.
- [5] T. Kohonen: Self-organization of topologically correct feature maps. Biological Cybernetics 43(1988). pp. 59–69.