Une nouvelle méthode pour l'extraction de paramètres : l'analyse en composantes curvilinéaires supervisée

Hicham Laanaya*,**, Arnaud Martin*
Ali Khenchaf* Driss Aboutajdine**

*ENSIETA-E³I²-EA3876, 2, rue François Verny 29806 Brest cedex 9,
laanayhi, Arnaud.Martin, Ali.Khenchaf@ensieta.fr
http://www.ensieta.fr/e3i2/
**GSCM-LRIT, Université Mohammed V-Agdal, Faculté des sciences de Rabat, Maroc
aboutaj@fsr.ac.ma,
http://www.fsr.ac.ma/GSCM/

Résumé. Nous présentons dans cette étude une nouvelle approche supervisée pour l'extraction de paramètres en se fondant sur l'analyse en composantes curvilinéaires. Cette approche prend en compte l'appartenance à des classes des données en vue de les classifier. Le principe de l'analyse en composantes curvilinéaires supervisée est de projeter les données dans un espace de petite dimension en gardant la topologie locale pour chaque classe. Nous évaluons cette approche après classification.
Mots clés : Extraction de paramètres, analyse en composantes curvilinéaires, analyse en composantes curvilinéaires supervisée.

1 Introduction

Les performances d'un classifieur dépendent du nombre des vecteurs utilisés pour leur apprentissage, de la dimension de ces vecteurs et de la complexité du classifieur. La réduction de la dimension permet d'alléger l'étape d'apprentissage du classifieur. Ils existent plusieurs méthodes dans la littérature pour pailler au problème de la dimension des données. Certaines approches tentent de conserver l'inertie telles que les analyses en composantes principales ou analyse factorielle discriminante qui en est sa version supervisée. Ces méthodes tentent de projeter les points sur un espace de dimension réduite qui est une transformation linéaire du premier espace. Des versions non-linéaire ont été proposée telle que l'analyse en composante indépendante. Une autre classe de méthodes tentent de conserver la topologie des données. Citons par exemple les algorithmes non-linéaires telles que Isometric Mapping Tenenbaum et al. (2000), Locally Linear Embedding Roweis et Saul (2000).

L'analyse en composantes curvilinéaires introduite dans Demartines et Hérault (1998) permet également de s'affranchir des hypothèses de linéarité en cherchant à conserver les topologies locales. Le fait de ne considérer que les topologies locales permet de traiter de très grande dimension. Dans une optique de classification de ces données nous avons introduit une version supervisée dans Laanaya et al. (2005). Nous donnons dans la section suivante un aperçu de la méthode de l'analyse en composantes curvilinéaires, nous présentons ensuite cette nouvelle approche supervisée de l'analyse en composants curvilinéaires. La dernière section présente quelques résultats de classification sur données générées avec l'analyse en composants curvilinéaires dans ces versions supervisée et non-supervisée.

2 Analyse en composantes curvilinéaires

L'analyse en composantes curvilinéaires introduite par Demartines et Hérault (1998), ou ACC, permet de s'affranchir des hypothèses de linéarité et de traiter des types de structures de données assez variées. La simplicité des calculs qu'elle met en jeu et son mode de convergence la rendent plus facile à utiliser que d'autres.

Cette approche est effectivement un outil puissant pour représenter des données d'un espace de très grande dimension dans un espace de petite dimension sans grande perte d'information locale.

2.1 Principe de l'analyse en composantes curvilinéaires

Le but de l'ACC est de reproduire la topologie de l'espace initial $\mathbb{R}^D$ dans un espace de dimension inférieure $\mathbb{R}^d$ dans lequel nous souhaitons projeter l'ensemble des observations. Il est difficile de conserver la topologie générale dans un espace de dimension inférieure. Nous essayons donc de reproduire la topologie locale. Le principe consiste alors à minimiser un critère caractérisant la différence de topologies entre l'espace initial et l'espace de projection, en favorisant les petites distances. Pour cela, posons :

$$E = \frac{1}{2} \sum_i \sum_{j \neq i} (X_{ij} - Y_{ij}) F_{\lambda(t)}(Y_{ij}), \tag{1}$$

où $X_{ij} = d(x_i, x_j)$, $Y_{ij} = d(y_i, y_j)$, avec x_i un vecteur de l'espace de départ et y_i son représentant dans l'espace d'arrivée, et $F_{\lambda(t)} : \mathbb{R}^+ \to [0,1]$ est une fonction décroissante visant à favoriser les petites distances. Habituellement $F_{\lambda(t)}$ est une fonction échelon : $F_\lambda(Y_{ij}) = I_{[0,\lambda(t)]}(Y_{ij})$.
La minimisation de E est faite en utilisant une méthode de gradient modifié :

$$\nabla_i E = \sum_{j \neq i} \frac{X_{ij} - Y_{ij}}{Y_{ij}} [2F_{\lambda(t)}(Y_{ij}) - (X_{ij} - Y_{ij}) F'_{\lambda(t)}(Y_{ij})](y_j - y_i), \tag{2}$$

où $\nabla_i E$ dénote le gradient de E par rapport à y_i.
Une minimisation de E par descente de gradient donne la règle d'adaptation suivante :

$$\Delta y_i = \alpha(t) \sum_{j \neq i} \frac{X_{ij} - Y_{ij}}{Y_{ij}} [2F_{\lambda(t)}(Y_{ij}) - (X_{ij} - Y_{ij}) F'_{\lambda(t)}(Y_{ij})](y_j - y_i). \tag{3}$$

Cette règle souffre de plusieurs défauts :
- Il faut faire, pour chaque i, une somme sur tous les j d'où une complexité de $O(N^2)$, avec N le nombre de vecteurs de l'espace de départ.
- Le processus peut tomber dans un minimum local de E.

Une procédure simple pour éviter ces défauts consiste à choisir un point y_i, d'une façon aléatoire, et tous les $y_{j\neq i}$ sont déplacés par rapport à lui. La nouvelle règle d'adaptation est donnée par :

$$\forall j \neq i \;\; \Delta y_j = \alpha(t) F_{\lambda(t)}(Y_{ij})(X_{ij} - Y_{ij}) \frac{y_j - y_i}{Y_{ij}}. \tag{4}$$

La complexité n'est ainsi que de $O(N)$.
L'algorithme 1 présente les différentes étapes de l'ACC.

Algorithm 1 Algorithme de l'ACC

Require: N le nombre d'itérations et n le nombre de vecteurs
  Initialisation aléatoire des points y_i
  for $t = 0$ to N **do**
    Evaluer $\alpha(t)$ et $\lambda(t)$
    for $i = 1$ to d **do**
      $\Delta y_j = \alpha(t) F_{\lambda(t)}(Y_{ij})(X_{ij} - Y_{ij}) \frac{y_j - y_i}{Y_{ij}}$ avec $j \neq i$
    end for
  end for

2.2 Qualité de la projection

Nous disposons de quelques outils pour évaluer la qualité d'une projection, ce qui permet également de comparer les méthodes de réduction de la dimensionnalité entre elles. Le premier est simplement la valeur de la fonction d'erreur ; le second est la comparaison de la distance sur l'espace de départ avec la distance sur l'espace de projection. Cette seconde technique est dénommée "représentation $\mathbf{dy} - \mathbf{dx}$"' Demartines et Hérault (1998) mais également "distorsion métrique".

Erreur de projection Le calcul des erreurs de projection, comme par exemple celle de l'ACC :

$$E_{ACC} = \frac{1}{2} \sum_i \sum_{j \neq i} (X_{ij} - Y_{ij}) F_\lambda(Y_{ij}) \tag{5}$$

ne permet pas d'évaluer directement la qualité d'une projection, car cette erreur dépend des données. En revanche, on peut utiliser cette erreur pour évaluer, sur un même jeu de données, la pertinence du choix des paramètres. Ainsi, on peut chercher des paramètres optimaux pour la méthode sur des données particulières.

Diagramme $\mathbf{dy} - \mathbf{dx}$ Demartines Demartines et Hérault (1998) propose de tracer le graphe des distances entre un certain nombre de points dans l'espace de départ en fonction des distances dans l'espace de représentation pour évaluer la qualité d'une projection. Si la topologie n'est pas respectée, les deux distances ne sont pas corrélées et on obtient un nuage diffus ; en revanche, une proportionnalité parfaite entre la distance dans l'espace de projection et celle de l'espace initial traduit un bon respect de la topologie initiale et est, globalement, le résultat d'une bonne projection.

3 Analyse en composantes curvilinéaires supervisée

Nous avons introduit la notion de l'analyse en composantes curvilinéaires pour traiter des données connaissant la classe de chaque individu. Pour cela, nous avons modifié l'algorithme de l'ACC pour prendre en compte la classe des individus, ainsi l'algorithme de l'ACC supervisée sera comme suit :

Soient $x_i \in \mathbb{R}^D,\ i = 1, ..., n$ les individus à représenter et $y_i \in \mathbb{R}^d,\ i = 1, ..., n$ les représentants des x_i dans le nouvel espace. Les étapes de l'ACC supervisée sont les mêmes que pour l'ACC, mais effectuées par classe. Elles sont décrites par l'algorithme 2.

Algorithm 2 Algorithme de l'ACC supervisée

Require: N le nombre d'itérations et n le nombre de vecteurs
  Initialisation aléatoire des y_i
  for $t = 0$ to N **do**
    Evaluer $\alpha(t)$ et $\lambda(t)$
    for $i = 1$ to d **do**
      $\Delta y_j = \alpha(t) F_{\lambda(t)}(Y_{ij})(X_{ij} - kY_{ij})\frac{y_j - y_i}{kY_{ij}}$ avec $j \neq i$ et Classe(x_i)=Classe(x_j)
    end for
  end for

Nous cherchons ainsi à conserver les topologies locales pour chaque classe prise indépendamment. Ainsi, pour chaque y_i, on déplace les y_j tel que $i \neq j$ et y_i et y_j sont de la même classe. Le k utilisé permet de contrôler le niveau de regroupement des individus de chaque classe ; il peut par exemple être donnée par l'inertie intra-classe. Ainsi un simple classifieur peut être utilisé pour séparer les classes facilement.

Pour l'étape d'affectation, le problème qui se pose est de représenter les nouveaux vecteurs, qu'on appelle points supplémentaires, qui ne font pas partis des x_i utilisés. Pour cela, étant donné un nouveau vecteur x_0, pour trouver le y_0 correspondant, il suffit de minimiser l'erreur suivante en utilisant la méthode de descente du gradient :

$$E_0 = \frac{1}{2} \sum_{i \neq 0} (X_{i0} - Y_{i0})^2 F_\lambda(Y_{i0}). \tag{6}$$

4 Expérimentations

Les expérimentations de la nouvelle approche ont été conduites sur des jeux de données gaussiennes générées de dimension 100 en 3 classes avec des moyennes et des variances différentes de façon à ce que les classes soient espacées. Nous avons utilisé des tirages aléatoires pour former les deux bases de données, une pour l'apprentissage et l'autre pour la validation de tailles 1000 vecteurs pour les deux bases de données.

Nous avons utilisé 10 itérations et la fonction exponentielle pour la fonction qui favorise les petites distances. Nous présentons ci-dessous les résultats obtenus en utilisant la méthode des Machines à Vecteurs de Support pour la classification Vapnik (1998). Nous avons utilisé le logiciel *libSVM* pour nos tests en utilisant un noyau linéaire et $C = 1$.

sans ACC supervisée avec ACC supervisée

$$\begin{pmatrix} 81.58 & 14.80 & 3.62 \\ 14.09 & 80.87 & 5.03 \\ 3.36 & 4.70 & 91.95 \end{pmatrix} \quad \begin{pmatrix} 90.38 & 8.59 & 1.03 \\ 9.00 & 87.33 & 3.67 \\ 3.24 & 4.21 & 92.56 \end{pmatrix}$$

Nous avons ainsi obtenu des taux de classifications respectifs de $84.77\pm2.23\%$ avant l'ACC supervisée et $90.11\pm1.85\%$ après l'application de l'ACC supervisée et des vecteurs d'erreurs de [13.57 14.44 6.19] sans application de l'ACC supervisée et l'ACC [7.87 9.53 4.90] avec une extraction par ACC supervisée.

5 Conclusion

Nous avons proposé dans ce papier une nouvelle approche pour l'extraction de paramètres fondée sur une version supervisée de l'analyse en composantes curvilinéaires. Cette version conserve toutes les propriétés de l'analyse en composantes curvilinéaires. Elle a de plus a montré son intérêt pour l'amélioration des performances de classification sur des données générées.

Références

Demartines, P. et J. Hérault (1998). Curvilinear component analysis : a self-organizing neural network for nonlinear mapping of data sets. *IEEE Transaction on Neural Networks 8(1)*, 711–720.

Laanaya, H., A. Martin, D. Aboutajdine, et A. Khenchaf (20-23 June 2005). A new dimensionality reduction method for seabed characterization : Supervised curvilinear component analysis. *IEEE OCEANS'05 EUROPE, Brest, France*.

Roweis, S. et L. Saul (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science 290(5500)*, 2323–2326.

Tenenbaum, J. B., V. de Silva, et J. C. Langford (2000). A global geometric framework for nonlinear dimensionality reduction. *Science 290*, 2319–2323.

Vapnik, V. N. (1998). *Statistical Learning Theory*. John Wesley and Sons.

Summary

In this study, we present a new supervised approach for feature extraction founded under curvilinear component analysis aproach. This approach take into account the data classes. The principle of the supervised curvilnnear component analysis is to proejct the data into a space with low dimension by preserving the local topology of each classe.
Keyords : Feature extraction, curvilinear component annalysis, supervised curvilinear component analysis.