A-4-21

オンラインK平面クラスタリングによる疎行列解析

Sparse Blind Identification by using Online K-plane Clustering

鷲沢 嘉一　　アンジェイ チホツキ
Yoshikazu Washizawa　　Andrzej Cichocki

(独) 理化学研究所 脳科学総合研究センター
RIKEN, Brain Science Institute

1 Introduction

We propose a blind source identification method under the assumption that source signals have many zero or negligibly small samples . Some kinds of signals are often sparse in either time domain, frequency domain or time-frequency domain. In order to identify unknown mixing matrix using this sparse property, we developed online learning for k-plane clustering. We provide an algorithm of this method and experimental results.

2 SCA by Alternating Projection

Let m, n and T be the number of observations, source signals and samples. A linear mixing model is defined as

$$\boldsymbol{X} = \boldsymbol{AS} + n, \tag{1}$$

where $\boldsymbol{X} \in \mathbb{R}^{m\times T}$ is a matrix of the observed signals, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a mixing matrix, $\boldsymbol{S} \in \mathbb{R}^{n\times T}$ is a matrix of source signals, and n is additional noise. $\boldsymbol{x}_t$ and $\boldsymbol{s}_t$ denote t-th column of $\boldsymbol{X}$ and $\boldsymbol{S}$, respectively.

If a time instant t, $\boldsymbol{s}_t$ has k zero components in $i_1, \ldots, i_k$ rows, then $\boldsymbol{x}_t$ is on a subspace spanned by column vectors of $\boldsymbol{A}$ except $i_1, \ldots, i_k$ columns. If $k + m > n$, $\boldsymbol{x}_t$ is on a proper subset (i.e., lower dimension) of $\mathbb{R}^m$.

Then, we define a criterion for space blind identification as

$$\hat{\boldsymbol{A}} = \underset{\boldsymbol{W}\in\mathbb{R}^{m\times n}}{\operatorname{argmax}} \sum_{t=1}^{T} \max_{\mathcal{I}_j\in\mathcal{I}} \|P_{\mathcal{R}(\boldsymbol{W}_{\mathcal{I}_j})}\boldsymbol{x}_t\|^2/|\mathcal{I}_j|^\gamma, \tag{2}$$

where $\mathcal{I}$ is a set of all subsets of $\{1, \ldots, n\}$, whose size is less than $m-1$, $\mathcal{I}_j$ is a element of $\mathcal{I}$ i.e., a subset of $\{1, \ldots, n\}$, $\boldsymbol{W}_{\mathcal{I}_j}$ is a matrix generated by column vectors of $\boldsymbol{W}$ with an index $\mathcal{I}_j$, $P_{\mathcal{R}(W_{\mathcal{I}_j})}$ is a projection matrix onto a range of $W_{\mathcal{I}_j}$, and $\gamma > 0$ is a parameter which control a sparsity for $\boldsymbol{S}$. For large γ, the more sparsity effects the smaller cost (for many outliers and strong sparseness), for small γ, the more samples lie on subspaces effects the smaller cost (for few outliers and weak sparseness).

Since it is difficult to get $\hat{A}$ directly, we obtain candidate vectors of $\hat{A}$ using following steps.

1. Remove column vectors from X that are close to the origin.
2. Normalize and whiten X. Whitening is not always necessary, but it improves condition number of A.
3. for $j \leftarrow 1$ to $\binom{n}{m-1}$
   (a) Initialize normal vector $\boldsymbol{w}_j$ randomly and normalize it to unit vector.
   for $t = 1, 2, \ldots,$learntime
      i. for $i = 1, 2, \ldots, T$
         - If $|\langle \boldsymbol{w}_j, \boldsymbol{x}_i\rangle| \le \theta(t)$, update $\boldsymbol{w}_j$
           $\boldsymbol{w}_j \leftarrow \boldsymbol{w}_j - \eta(t)\langle \boldsymbol{w}_j, \boldsymbol{x}_i\rangle \boldsymbol{x}_i$,
           $\boldsymbol{w}_j \leftarrow \boldsymbol{w}_j/\|\boldsymbol{w}_j\|$
   (b) Remove samples such that $d_i \le \theta(t)$
4. Obtain normal vectors of hyperplanes spanned by every $(m-1)$ vectors which are subset of a set of $\boldsymbol{w}_j$.
5. Obtain $\hat{A}$ using a criterion (2).

$\theta(t)$ and $\eta(t)$ are threshold and learning coefficient which decrease suitably during learning.

If a column of $\boldsymbol{S}$ has more than $(n-m+1)$ zero components in most of instants, observed vector lies on a hyperplane generated from columns of A. A lot of columns which have more than $(n-m+1)$ zero components focus on $\binom{n}{m-1}$ $(m-1)$-dimantional hyperplane and 1-dimensional subspace which are generated by intersections of the subspaces correspond cokumns of $\boldsymbol{A}$.

3 Computational Simulation

We generated random five source signals which have 80% zero components in each row and a mixing matrix as

$$\boldsymbol{A} = \begin{bmatrix} 0.8412 & 0.0298 & 0.0750 & 0.1735 & 0.7240 \\ -0.5025 & -0.8305 & 0.8294 & -0.3621 & 0.4088 \\ 0.1997 & -0.5563 & -0.5536 & -0.9158 & -0.5556 \end{bmatrix}.$$

Each column of $\boldsymbol{A}$ is normalized. Estimated mixing matrix is

$$\hat{A} = \begin{bmatrix} 0.0730 & 0.0299 & 0.8419 & 0.7242 & 0.1739 \\ 0.8302 & -0.8302 & -0.5016 & 0.4088 & -0.3606 \\ -0.5526 & -0.5566 & 0.1990 & -0.5554 & -0.9164 \end{bmatrix}.$$

4 Conclusion

We proposed new blind identification method by using alternating projection. The computational simulations confirm that it is possible to recover a mixing matrix under overcompleate case $(m < n)$.