УДК 681.142.2:519.24 *ИНФОРМАТИКА*

МЕТОД АНАЛИЗА НЕЗАВИСИМЫХ КОМПОНЕНТ НА ОСНОВЕ ПРЕОБРАЗОВАНИЯ НЕЗАВИСИМОСТИ

Горячкин О.В., Шатских С.Я.

Представлено академиком РАН Ю.В. Прохоровым

Введение. Одна из центральных проблем в практике приложений нейронных сетей, статистике, задачах цифровой обработки сигналов и изображений - это задача нахождения наиболее компактного представления данных. Это важно для последующего анализа, которым может быть распознавание образов, классификация и принятие решений, сжатие данных, фильтрация шумов, визуализация. Относительно недавно, для решения подобных задач, привлек широкое внимание метод нахождения преобразования, обеспечивающего независимость компонент наблюдаемого вектора данных, названный в [1] анализом независимых компонент (АНК). Модель, используемую в анализе независимых компонент, можно представить в виде

$$\mathbf{y} = \mathbf{H}\mathbf{x}, \tag{1}$$

где: $\mathbf{y}$ - m-мерный случайный вектор, $\mathbf{x}$ - n-мерный случайный вектор с независимыми компонентами, $\mathbf{H}$ - некоторое обратимое неизвестное отображение $R^n \to R^m$, $m \geq n$. Задача АНК формулируется как задача поиска такой проекции вектора $\mathbf{y}$ на линейное пространство векторов $\mathbf{x}$, компоненты которой были бы статистически независимы. При этом для анализа доступна только некоторая статистическая выборка значений случайного вектора $\mathbf{y}$. В линейном анализе независимых компонент, когда $\mathbf{H}$ это детерминированная, неизвестная $m \times n$ матрица, АНК является развитием хорошо известных в прикладной статистике методов анализа принципиальных компонент и факторного анализа, где вместо свойства некоррелированности используется более сильное свойство статистической независимости. Традиционные методы линейного АНК строятся по вариационному принципу [2]

$$\mathbf{x} = \arg \min_{\mathbf{A}} \vee \max \left(Q(\mathbf{Ay}) \right), \tag{2}$$

где $\mathbf{A}$ - $n \times m$ матрица, Q - функционал, имеющий смысл критерия независимости компонент. Для нелинейной модели АНК, задача как правило является недоопределенной, поскольку часто неясен вид отображения $\mathbf{H}$ в (1). Соблазнительным решением проблемы АНК в этом случае было бы явное определение преобразования независимости компонент [3].

Преобразование независимости. Рассмотрим случай, когда случайные вектора $\mathbf{x}$ и $\mathbf{y}$ имеют совместные функции распределения компонент, которые вместе со всеми своими маргинальными распределениями непрерывны и всюду положительны. Рассмотрим отображение $H^{-1} : R^n \to R^n$ с координатными функциями вида [3]

$$\begin{cases} x_n = F_n^{-1}\left(\left(F_{n|1\ldots n-1}(y_n \mid y_1, \ldots, y_{n-1})\right)\right), \\ \vdots \\ x_k = F_k^{-1}\left(F_{k|1\ldots k-1}(y_k \mid y_1, \ldots, y_{k-1})\right), \\ \vdots \\ x_1 = y_1 \end{cases} \tag{3}$$

где $F_{k|1...k-1}(y_k \mid y_1,...,y_{k-1})$ - условная функция распределения случайной величины y_k, $F_k^{-1}(\bullet)$ - обратная функция, соответствующая одномерной функции распределения случайной величины y_k. Как было показано в [3] система случайных величин $\{x_k\}_{k=1,...,n}$ взаимно независима. Введенное в [3] треугольное преобразование независимости (3) фактически является вариантом преобразования введенного в работе [5]. Хорошо известно, что для гауссовских случайных векторов свойство независимости их компонент эквивалентно их некоррелированности. Поэтому для каждого гауссовского вектора существует линейное преобразование ортогонализации, совпадающее с (3). В случае произвольного распределения компонент $\{y_k\}_{k=1,...,n}$ преобразование (3) нелинейное.

Алгоритм АНК. В задаче АНК мы не имеем информации о функции распределения $F_n(y_1,...,y_n)$, поскольку нам доступна только выборка значений случайного вектора $\mathbf{y}$. Поэтому рассмотрим путь построения (3) по выборочным статистикам. Т.к. для построения преобразования независимости требуется непрерывность $\hat{F}_n(y_1,...,y_n)$, а также всех маргинальных распределений, то целесообразно использовать ядерные оценки функций распределения в виде

$$\hat{F}_n(y_1,...,y_n) = \frac{1}{N^n}\sum_{i_1=1}^{N}...\sum_{i_n=1}^{N}\chi(y_1 - y_1(i_1))...\chi(y_n - y_n(i_n)), \qquad (4)$$

где $\chi(y) = \int\limits_{-\infty}^{y}\mu(z)dz$, $\mu(x)$ - положительная функция окна, такая что $\int \mu(x)dx = 1$. Если n велико (на практике более 3-х), то в этом случае трудно получить достоверные оценки многомерных распределений с достаточной точностью. Возможность построения преобразования независимости n-мерного случайного вектора с помощью парных преобразований независимости для негауссовских случайных векторов была найдена в [4]. В [4] показано, что достаточным условием возможности построения преобразования независимости n-мерного случайного вектора с помощью парных преобразований независимости является свойство воспроизводимости условных квантилей многомерного распределения. Условные квантили $q^0_{i|1\ldots\tilde{i}\ldots N}(y_1,\ldots,\tilde{y}_i,\ldots y_N)$ распределений $F_{k|1\ldots k-1}(y_k \mid y_1,\ldots,y_{k-1})$ определим следующими уравнениями:

$$\begin{aligned} &F_{i|1\ldots\tilde{i}\ldots N}\left(q^0_{i|1\ldots\tilde{i}\ldots N}(y_1,\ldots,\tilde{y}_i,\ldots y_N) \mid y_1,\ldots,\tilde{y}_i,\ldots,y_N\right)= \\ &= F_{i|1\ldots\tilde{i}\ldots N}\left(y_i^0 \mid y_1^0,\ldots,\tilde{y}_i^0,\ldots,y_N^0\right) \end{aligned}, \tag{5}$$

где символ “~” над переменной означает ее исключение. В соответствии с [4] случайный вектор обладает свойством воспроизводимости условных квантилей размерности n-1 при сужении на одномерные условные квантили, если для любого $i = 1,\ldots,n$ и для любого $k = 1,\ldots,n$ такого, что для $k \neq i$

$$q^0_{i|1\ldots\tilde{i}\ldots N}\left(q^0_{1|k}(y_k)\ldots q^0_{k-1|k}(y_k), y_k, q^0_{k+1|k}(y_k)\ldots\tilde{x}_i\ldots q^0_{N|k}(y_k)\right) = q^0_{i|k}(y_k) \tag{6}$$

В работе [4] приведены примеры многомерных распределений, условные квантили которых обладают свойством воспроизводимости. Это распределения Гаусса, Стьюдента, Коши, Дирихле и некоторые типы сопряженных распределений. Можно показать, что этим свойством обладает распределение случайного вектора полученного с помощью линейного однозначного отображения вектора с независимыми, произвольно, но одинаково распределенными компонентами. Процедура построения преобразования независимости n-мерного случайного вектора с помощью парных преобразований независимости может быть сведена к следующим этапам:

1) Пусть мы имеем набор реализаций n случайных величин $\{y_k\}_{k=1,...,n}$. По набору реализаций построим n-1 выборочных условных распределений $\hat{F}^1_{k|1}(y_k \mid y_1)$, k>1. Получим набор реализаций n-1 случайных величин $\{y^1_m\}_{m=1,...,n-1}$ используя преобразование $y^1_m = \hat{F}^1_{m+1|1}(y_{m+1} \mid y_1)$.
2) По набору реализаций $\{y^1_k\}_{k=1,...,n-1}$ построим n-2 выборочных условных распределений $\hat{F}^2_{k|1}(y^1_k \mid y^1_1)$, k>1. Получим набор реализаций n-2 случайных величин $\{y^1_m\}_{m=1,...,n-2}$ используя преобразование $y^2_m = \hat{F}^2_{m+1|1}(y^1_{m+1} \mid y^1_1)$.
3) Продолжая этот процесс, получим набор реализаций случайной величины y^{n-1}_1 и соответствующую предыдущему этапу выборочную функцию распределения $\hat{F}^{n-1}_{2|1}(y^{n-2}_2 \mid y^{n-2}_1)$.
4) Используя полученный набор двумерных выборочных условных функций распределения преобразование независимости может быть построено как рекуррентная система равенств (7).

$$\begin{cases} x_1' = y_1, \\ x_2' = \hat{F}_1^{-1}\left(\hat{F}_{2|1}^{1}\left(y_2 \mid x_1'\right)\right), \\ x_3' = \hat{F}_1^{-1}\left(\hat{F}_{2|1}^{2}\left(y_2^1 \mid \hat{F}_1^{-1}\left(x_2'\right)\right)\right), \\ \ldots\ldots\ldots\ldots\ldots \\ x_n' = \hat{F}_1^{-1}\left(\hat{F}_{2|1}^{n-1}\left(y_2^{n-2} \mid \hat{F}_1^{-1}\left(x_{n-1}'\right)\right)\right). \end{cases} \tag{7}$$

Предложенный метод АНК может быть использован в задачах слепой идентификации и коррекции скалярных или векторных каналов, а также многоканальных систем [7]. В частности, дало положительные результаты применение преобразования (3) при решении задачи компенсации зависимости между изображениями различных спектральных зон при совместной обработке многозональных оптических космических изображений [6].

Выводы. Предложенный алгоритм обеспечивает возможность проведения АНК по наблюдаемой выборке путем построения координатных функции преобразования независимости в явном виде. В частном случае, при удовлетворении распределения данных свойству воспроизводимости условных квантилей, преобразование независимости может быть построено только по двумерным выборочным распределениям.

Список литературы

1. Comon P. // Signal Processing. 1994. Vol. SP-36. P. 287-314.
2. Hyvarinen A. // Neural computing surveys. 1999. N2. P.94-128.
3. Шатских С.Я. // Сб. “Мера и интеграл”. СГУ. 1995. С.99-112.
4. Шатских С.Я. // Вестник Самарского государственного университета: естественнонаучная серия. №1. 2002. С.36-58.
5. Rosenblatt M. // Ann. Math. Stat. N23. 1952. P.470-472.
6. Goriachkin O.V., Filimonov A.R., Klovsky D.D., Shatskih S.J. // Proc. of Third International Airborne Remote Sensing Conference and Exhibition. 1997. Vol.1. P.387-392.
7. Горячкин О.В. Методы слепой обработки сигналов и их приложения в системах радиотехники и связи. М.: Радио и связь. 2003. С.230.

Горячкин Олег Валериевич,

Поволжская государственная академия телекоммуникаций и информатики,

443010 г. Самара, ул. Льва Толстого 23, ПГАТИ, кафедра ТОРС,

(8462) 391143 (служ.), (8462) 354543 (дом.), (8462) 335856 (факс),

E-mail: gor@mail.radiant.ru

Шатских Сергей Яковлевич,

Самарский государственный университет,

443011 г. Самара, ул. Акад. Павлова 1, кафедра ТВиМС.

(8462) 780937 (служ.), (8462) 428655 (дом.),

E-mail: shatskih@ssu.samara.ru