

Introducción a la separación ciega de fuentes

Por Juan Manuel Miguel Jiménez, Luciano Boquete Vázquez

Departamento de
Electrónica. Universidad
de Alcalá de Henares.
E-mail:
jmanuel@depeca.uah.es,
boquete@depeca.uah.es

El problema de la separación ciega de fuentes consiste en la obtención de cada una de las señales generadas por n fuentes estadísticamente independientes, a partir de las mezclas lineales que tienen lugar en el propio medio en el que se propagan y que son detectadas al menos por n sensores. Diversos algoritmos desarrollados en los últimos años, hacen de esta técnica una herramienta prometedora en un sinfín de aplicaciones, entre las que han comenzado a obtener fruto el filtrado de ruido, procesamiento de señales de audio, señales astrofísicas, electrocardiografía fetal, electroencefalografía y el análisis y mejora de imágenes médicas.

Uno de los problemas en electromedicina, tras el registro de las señales capturadas por los sensores, es el complicado análisis de las mismas en busca de patrones y parámetros que reflejen patologías concretas, todo ello por la gran cantidad de información, muchas veces de poco interés y otras, redundante y mezclada con diversas fuentes de ruido, que ocultan las estructuras esenciales de las señales buscadas.

Otro ejemplo típico de separación ciega de fuentes (BSS, Blind Source Separation) es el problema conocido como cocktail party. En este caso, se colocan varios micrófonos que registran una mezcla de varias señales acústicas que se producen de manera simultánea. El propósito es separar los sonidos y voces originales de cada uno de los hablantes, y para ello es necesario colocar al menos tantos micrófonos como parlantes.

Existe otro gran número de aplicaciones que contienen datos o señales multidimensionales, donde la separación ciega de fuentes encuentra una aplicación ideal, como son señales e imágenes biomédicas (por ejemplo, separación de las señales electrocardiográficas materna y del feto), misiones multisatélite (análisis de observaciones que proceden de diver-

sos satélites), separación de señales en telecomunicaciones, búsqueda de factores ocultos en mercados financieros, etc.

El problema del Cocktail Party

Ilustremos mediante el mencionado *cocktail party* el proceso global de la separación de señales. Supongamos que nos encontramos en una sala donde hay dos personas que hablan simultáneamente, y tenemos dos micrófonos colocados en diferentes lugares que registran las respectivas señales variables en el tiempo, $x_1(t)$ y $x_2(t)$. Si consideramos que las señales emitidas por cada uno de los parlantes son $s_1(t)$ y $s_2(t)$, respectivamente, las señales recogidas por cada uno de los micrófonos será una suma ponderada de estas últimas y podremos expresarlas mediante un par de ecuaciones lineales:

$$\begin{aligned} x_1(t) &= m_{11}s_1 + m_{12}s_2 \\ x_2(t) &= m_{21}s_1 + m_{22}s_2 \end{aligned} \quad (1)$$

donde m_{ij} son parámetros que dependen principalmente de la distancia de cada micrófono a los respectivos parlantes. En notación matricial, podemos expresar esto mismo como:

$$X = MS \quad (2)$$

siendo:

$$\begin{aligned} X &= [x_1(t), x_2(t)]^T \\ S &= [s_1(t), s_2(t)]^T \end{aligned}$$

denotando por por T la traspuesta o vector columna,

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$$

A la matriz M se le conoce como matriz de mezcla. Es inmediato pensar que para recuperar las señales fuente originales, bastaría con calcular la inversa de la matriz de mezcla y multiplicarla por las señales

mezcladas. Lo espectacular de la BSS es obtener las señales originales partiendo únicamente de las señales mezcladas X como datos, pero sin conocer la matriz de mezcla, de ahí el adjetivo "ciega". Por tanto, debemos estimar los coeficientes de la matriz M y calcular su inversa, la matriz W , llamada *matriz de separación*, que nos permita recuperar lo más fielmente posible las señales fuente originales al multiplicarla por la matriz de señales mezcla. Las señales recuperadas, $y_1(t)$ y $y_2(t)$, deberían parecerse lo máximo posible a las señales originales y una vez calculada la matriz de mezcla serán obtenidas así:

$$Y = W \cdot X \quad (3)$$

Pero el problema de encontrar el valor de la matriz de separación partiendo exclusivamente de las señales mezcladas no es nada fácil y se han ideado varias técnicas, todas basadas en asumir que las señales fuente originales $s_1(t)$ y $s_2(t)$ son estadísticamente independientes en cada instante t . La técnica reciente que ha resultado ser más efectiva es el Análisis de Componentes Independientes (ICA, Independent Component Analysis).

No obstante, antes de adentrarnos en los detalles de ICA se propone un posible algoritmo poco práctico, pero muy intuitivo, para la separación ciega de fuentes y que podría consistir en encontrar unos valores de k_1 y k_2 tales, que se consiga máxima independencia entre las dos señales estimadas siguientes:

$$\begin{aligned} y_1 &= x_1 + k_1 x_2 \\ y_2 &= x_1 + k_2 x_2 \end{aligned} \quad (4)$$

Un doble bucle que recorriese un rango de valores para cada constante k_1 y k_2 , y que comprobase tal condición de independencia a partir de ciertos parámetros estadísticos, podría implementarlo. Así, para el siguiente ejemplo de señales mezcla:

$$x_1 = 3 \cdot s_1 + s_2$$

$$x_2 = s_1 + 2 \cdot s_2$$

la solución de k_1 y k_2 sería la que proporcionara máxima independencia entre y_1 e y_2 , es decir, que estas últimas contuviesen la información respectiva de las señales fuente originales, pero sin mezclar (cuadro 1):

hacer separación de sonidos mediante ICA” y “Esto es una divertida forma de entretenernos”, respectivamente. Las señales x_1 a x_4 representan las mezclas de las señales originales, con distintas ponderaciones, y que se presentan como entrada al algoritmo FastICA. Podemos observar que el número de mezclas de entra-

que se corresponden con s_3 , s_1 y s_2 , respectivamente. La calidad de la separación, si observamos dichas señales, es más que satisfactoria.

Asunciones previas y limitaciones del algoritmo ICA

Tal y como ya hemos comentado, la asunción clave de partida es suponer que las señales fuente originales son estadísticamente independientes. También, se supone que sus distribuciones son no gaussianas, aunque no necesariamente conocidas a priori ya que, de conocerlas, el problema sería mucho más sencillo. Otro aspecto a tener en cuenta es que ICA no sabe recuperar la componente continua de las señales originales, ni tampoco la polaridad de su componente alterna, aunque esto en la práctica no suele acarrear ningún problema. Por ello, algunas señales aparecerán invertidas 180 grados y otras con la fase correcta. Finalmente, tampoco se podrá determinar la amplitud original de las componentes independientes, y siempre se asumirá que tienen como valor de varianza la unidad.

Esta última ambigüedad encuentra su explicación en que, al ser desconocidas tanto las señales como la matriz de mezcla (ver ecuaciones 1 y 2), si multiplicamos cualquier señal fuente s_i por una constante, ésta puede ser cancelada al dividir los coeficientes de la correspondiente columna de la matriz de mezcla por la mencionada constante. Así, la interpretación física de ello es que obtenemos la misma magnitud en las señales mezcla o señales capturadas si multiplicamos por dos la amplitud de la señal s_1 , a que en su lugar hubiésemos multiplicado por dos a m_{11} y m_{21} forzando a decir en este último caso que es el medio por el que se propaga s_1 el que amplifica por dos, o atenúa la mitad, según se interprete.

Cuadro 1.

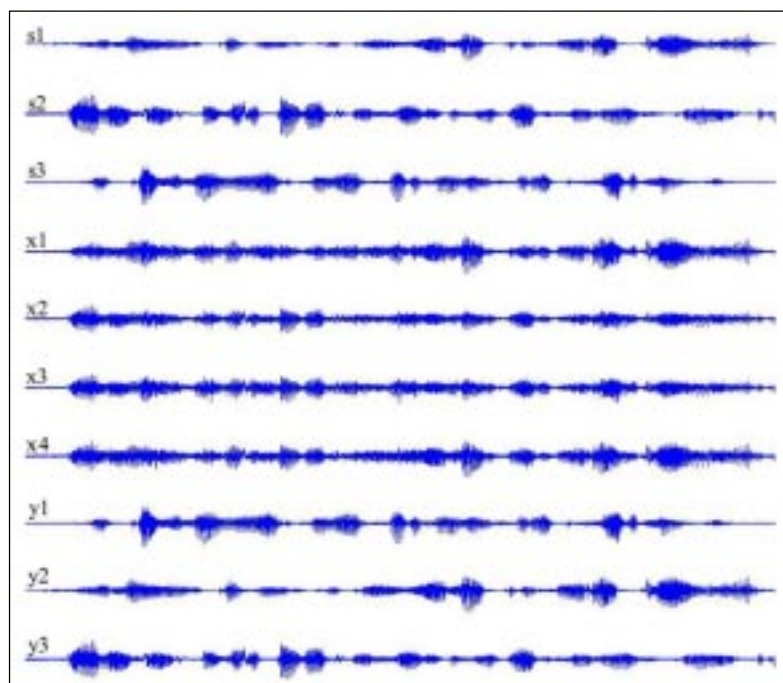
$$y_1 = 3s_1 + s_2 + k_1(s_1 + 2s_2) \rightarrow k_1 = -0.5 \rightarrow y_1 = 3s_1 + s_2 - 0.5s_1 - s_2 = 3.5s_1$$

$$y_2 = 3s_1 + s_2 + k_2(s_1 + 2s_2) \rightarrow k_2 = -3 \rightarrow y_2 = 3s_1 + s_2 - 3s_1 - 6s_2 = -5s_2$$

En separación ciega de fuentes, el algoritmo FastICA representa actualmente el estado del arte y, como la mayoría de algoritmos destinados a este fin, es implementado mediante redes neuronales. Un ejemplo de los resultados obtenidos al separar tres señales de audio se muestra en la figura 1.

da no tiene por qué ser idéntico al de señales independientes, siendo superior en este caso.

Antes de proceder a la separación de las señales, FastICA realiza, entre otras operaciones que forman parte del preprocesado, un análisis de componentes principales (PCA), con el que determina el número más



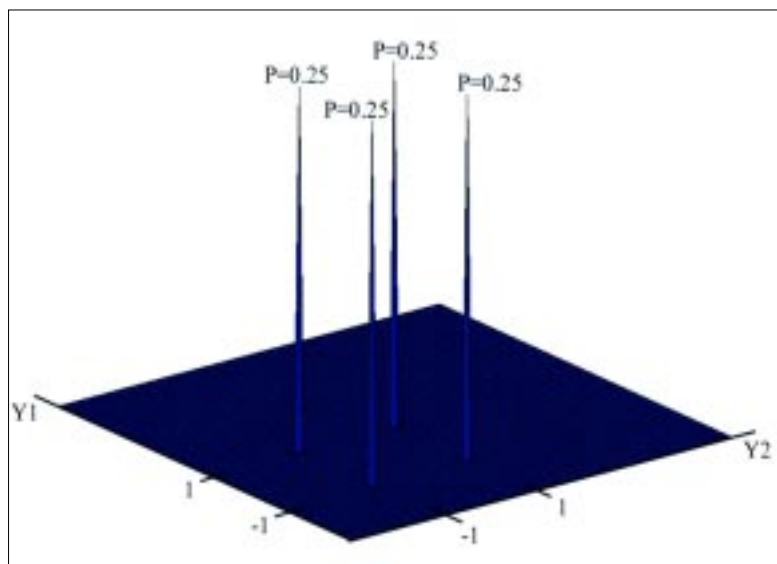
En dicha figura, s_1 , s_2 y s_3 representan las señales originales de audio que se grabaron y que corresponden a las locuciones: “Hola, esto constituye la primera prueba”, “Vamos a

probable de componentes o señales independientes que componen todas las mezclas, en nuestro ejemplo tres.

Una vez efectuada la separación, obtenemos las salidas y_1 , y_2 e y_3 ,

Figura1. Ejemplo de separación de señales de audio.

Figura 2. Función densidad de probabilidad conjunta $p(y_1, y_2)$ del ejemplo.



La limitación de que las señales a separar no sean gaussianas es realmente la más importante y encuentra su explicación en la propiedad en la que nos basamos para separarlas. A partir del teorema del límite central en la teoría de probabilidades, sabemos que la distribución correspondiente a una suma de variables independientes tiende a ser gaussiana, aunque las distribuciones parciales no lo sean. Pues bien, esta es precisamente la propiedad más común en la que se basa el algoritmo ICA para separar las señales fuente que se supone son independientes y con distribuciones no gaussianas, a partir de las señales mezcla cuyas distribuciones son más o menos gaussianas, según el número de señales sumadas. Ajustando los coeficientes de la matriz de separación, de modo que se minimice la gaussianidad en la distribución de las señales resultantes, conseguiremos que las señales de salida sean lo más independientes posible, es decir, no exista información mutua entre ellas, asemejándose así a las señales fuente originales. En realidad, si de todas las componentes que forman las mezclas, sea cual sea su número, sólo es una la que tiene distribución gaussiana, el algoritmo ICA sigue trabajando bien.

Medidas de independencia

Dos variables s_1 y s_2 se dice que son independientes si el valor de la variable s_1 no aporta ninguna información sobre el posible valor de la variable s_2 , y viceversa. Esto, efectivamente es cierto en nuestras señales originales, pero no así en sus mezclas x_1 y x_2 ya que la una contiene información sobre la otra. Siempre tendemos a pensar en estadística, que un primer indicio en la sospecha de que dos variables aleatorias sean independientes es que sean no correladas, en cuyo caso su covarianza es cero. Pero una medida estricta de independencia entre ellas, no acepta

únicamente que sean decorreladas, sino que su función densidad de probabilidad conjunta sea factorizable y, como consecuencia de ello, si suponemos dos funciones cualesquiera de dichas variables $f_1(s_1)$ y $f_2(s_2)$, se podría demostrar que su covarianza siempre será cero, independientemente de las funciones tomadas:

$$E\{f_1(s_1)f_2(s_2)\} = E\{f_1(s_1)\}E\{f_2(s_2)\} = 0$$

Según esto, para mostrar un ejemplo de dos variables que son decorreladas, pero con ello no independientes, supongamos que tienen valores discretos y que su función densidad de probabilidad es tal que la

probabilidad siempre es 0.25 en los siguientes pares de valores (y_1, y_2) : $(0,1)$, $(0,-1)$, $(1,0)$, $(-1,0)$. A partir de esto, podemos representar su función densidad de probabilidad conjunta, como aparece en la figura 2.

Podemos comprobar que se trata de dos variables no correladas al resultar ser cero su covarianza (ver cuadro 2).

En cambio, ello no significa que sean independientes si tomamos:

$$f_1(y_1) = (y_1)^2 \text{ y } f_2(y_2) = (y_2)^2$$

como se demuestra a continuación (ver cuadro 3):

Cuadro 2

$$E\{y_1 y_2\} = E\{y_1\}E\{y_2\} = (0 \cdot 1 + 0 \cdot (-1) + 1 \cdot 0 + (-1) \cdot 0) \cdot 0.25 = (-1 + 1) \cdot 0.25 = 0$$

Cuadro 3

$$p(y_1) = \int p(y_1, y_2) dy_2 \Rightarrow \begin{cases} p(y_1 = -1) = 0.25 \\ p(y_1 = 0) = 0.5 \\ p(y_1 = 1) = 0.25 \end{cases}$$

$$E\{y_1^2\} = (-1)^2 \cdot 0.25 + (0)^2 \cdot 0.5 + (1)^2 \cdot 0.25 = 0.5$$

$$E\{y_2^2\} = 0.5$$

$$\Rightarrow E\{y_1^2\}E\{y_2^2\} = 0.25$$

$$E\{y_1^2 y_2^2\} = [(-1)^2 \cdot (0)^2] \cdot 0.25 + [(1)^2 \cdot (0)^2] \cdot 0.25 + [(0)^2 \cdot (-1)^2] \cdot 0.25 + [(0)^2 \cdot (1)^2] \cdot 0.25 = 0$$

De dicho cuadro se tiene:

$$E\{y_1^2 y_2^2\} = 0 \neq 0.25 = E\{y_1^2\} E\{y_2^2\} \quad (7)$$

lo que demuestra que y_1 e y_2 , a pesar de ser decorreladas, no son independientes.

Como ya se comentó anteriormente, al sumar dos o más variable aleatorias (para nosotros representan a las señales originales desconocidas, y que tenemos que separar), la distribución de la suma es más gaussiana que la de cualquiera de las componentes que se suman. Necesitamos, por tanto, una medida que estime la no gaussianidad, para que el algoritmo pueda ajustar los coeficientes de la matriz de separación de una manera eficiente. Varios estimadores, a veces llamados funciones objetivo, son los que se han propuesto para dicha medida, como la kurtosis y la negentropía. La kurtosis, o cumulante de cuarto orden, de una variable aleatoria x que ha sido previamente normalizada para que su varianza valga la unidad, es simplemente una versión normalizada del momento de cuarto orden:

$$kurt(x) = E\{x^4\} - 3 \quad (6)$$

La kurtosis de una variable aleatoria gaussiana será cero, mientras que para casi todas las que no sean gaussianas valdrá distinto de cero, salvo en muy raras excepciones. Por tanto, el algoritmo podría ser implementado en una red neuronal, de modo que esta ajustara los coeficientes de la matriz de separación, hasta encontrar un máximo en el valor de la kurtosis de las respectivas señales separadas. En la práctica, la kurtosis presenta algunos inconvenientes, resultando como medida más robusta de no gaussianidad la negentropía, obtenida a partir de la entropía. Esta última, según la teoría de la información, mide la cantidad de información contenida en una señal, siendo mayor la entropía de una señal cuanto más aleatoria sea esta. La entropía de una señal discreta x se define como:

$$H(x) = -\sum_i p(x=a_i) \log p(x=a_i) \quad (7)$$

donde $p(x=a_i)$ es la probabilidad de que la señal x tome el valor a_i .

Según la teoría de la información, una variable con distribución gaussiana tiene mayor entropía que si tuviese cualquier otro tipo de distribución, para un mismo valor de varianza. Según esto, podemos decir que la distribución gaussiana es la distribución más aleatoria que existe. Además, esto nos sirve para utilizar la entropía como medida de no gaussianidad, o bien una versión modificada de la misma que se llama negentropía y vale cero para el caso de una variable gaussiana:

$$J(x) = H(x_{\text{gausiana}}) - H(x) \quad (8)$$

donde $H(x_{\text{gausiana}})$ es la entropía de una variable aleatoria gaussiana, x_{gausiana} , cuya matriz de covarianza es la misma que la de x . En la práctica, se utiliza una aproximación de la expresión (8) que resulta computacionalmente mucho menos costosa.

El Algoritmo Infomax

Uno de los métodos más comunes para implementar ICA es el algoritmo *Infomax*, basado en la medida de la información mutua entre las componentes estimadas. La información mutua entre las componentes de un vector Y se define como:

$$I(Y) = \sum_i H(y_i) - H(Y) \quad (9)$$

donde $H(Y)$ denota la entropía obtenida de la densidad conjunta de probabilidad $f(Y)$ de todas las componentes y_i :

$$H(Y) = -\int f(Y) \log f(Y) dY \quad (10)$$

El algoritmo *Infomax* se lleva a cabo mediante una red neuronal cuya estructura general se muestra en la figura 3. Las señales de entrada x_i representan las mezclas

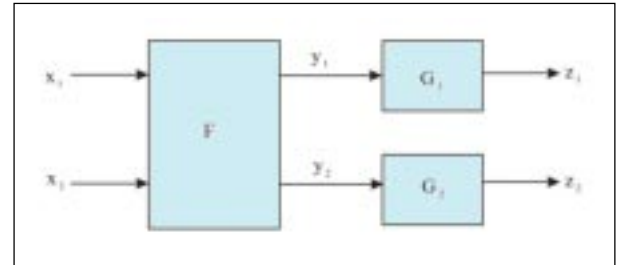


Figura 3. Diagrama de la red que implementa el algoritmo Infomax

capturadas y, por tanto, no son estadísticamente independientes. Se trata de que las señales y_i obtenidas sean tan independientes como sea posible. Para el caso más corriente de ICA lineal como el que nos ocupa, donde las señales de entrada capturadas son una mezcla lineal de las señales independientes a separar, la función F es una función de transformación lineal efectuando sencillamente un producto por una matriz, mientras que las funciones G_i son funciones no lineales.

Conocemos, a partir de la teoría de la información, que dos variables que estén relacionadas por una función invertible, sea o no lineal, contienen la misma cantidad de información.

Del mismo modo, la información mutua entre las componentes del vector Y en la figura 3, es igual a la información mutua entre las componentes del vector Z . Si, por otra parte, tomamos como funciones G_i la función de probabilidad acumulativa de la correspondiente componente y_i (algo parecido a cuando deseamos obtener una ecualización uniforme del histograma), entonces las componentes z_i tendrán una distribución uniforme y las entropías $H(z_i)$ valdrán cero, resultando:

$$I(Y) = I(Z) = \sum_i H(z_i) - H(Z) = -H(Z) \quad (11)$$

reduciéndose el problema de la minimización de información mutua entre componentes a una maximización de la entropía de salida $H(Z)$.

Bibliografía

- Aapo Hyvärinen and Erkki Oja. "Independent Component Analysis: Algorithms and Applications". Neural Networks. 1999.

- Luís B. Almeida. "MISEP – Linear and Nonlinear ICA Based on Mutual Information". Journal of Machine Learning Research. 2002.

- Andreas Jung. "An Introduction to a new data analysis tool: Independent Component Analysis". Proceedings of Workshop GK "Nonlinearity". 2002.

- The FastICA MATLAB package. <http://www.cis.hut.fi/projects/ica/fastica/>.

Aplicaciones de ICA

Las aplicaciones más inmediatas de ICA se encuentran en muchas áreas del procesamiento de señales, como el procesamiento de señales de audio, y en el procesamiento de imágenes donde ofrece alternativas a la reducción del ruido, separación de datos procedentes de múltiples satélites, etc. La figura 4 muestra un ejemplo bastante espectacular de separación de imágenes, previamente mezcladas. En la fila superior aparecen las imágenes originales de partida que, una vez mezcladas con distintos coeficientes de ponderación, dan lugar a las imágenes de la fila central, y que fueron los datos de partida para el algoritmo FastICA. El resultado es el conjunto de imágenes de la fila inferior, donde se aprecia la notable calidad de la separación, incluso sin haber

hecho un ajuste óptimo de los parámetros con los que trabaja dicho algoritmo. Merece la pena observar, que ICA no sabe presentar las salidas en el orden deseado.

En el campo de imágenes médicas, ICA se presenta como una herramienta revolucionaria para el análisis de datos biomédicos. Así, en la detección de regiones activas en el cerebro mediante imágenes de resonancia magnética, la separación ciega de fuentes ofrece grandes posibilidades. Otro ejemplo, es la extracción del electrocardiograma fetal (FECG) mediante técnicas no invasivas, a partir de las señales procedentes de electrodos colocados en la piel de la madre. En este caso, es necesario separar la débil señal fetal recogida por los electrodos externos, mezclada con fuertes perturbaciones como las propias señales de electro-

cardiograma (ECG) y electromiograma (EMG) materno, interferencias de la red, ruido térmico, etc. En electroencefalografía (EEG) y en la moderna magnetoencefalografía (MEG) donde se mide la actividad cortical de las neuronas, nos encontramos con multitud de señales entremezcladas que hacen muy difícil la obtención de características y datos esenciales por parte del investigador. Tales señales, van desde la superposición del gran número de éstas que se generan en diferentes zonas del cerebro, a los artefactos producidos durante el parpadeo de los ojos, movimientos de la cabeza, pulso del corazón, etc. Aquí, se hace necesario una separación en componentes independientes, para centrar el estudio sólo en aquellas señales que sean relevantes para el caso tratado. □



Figura 4. Ejemplo de separación de imágenes con FastICA. Fila superior: imágenes originales. Fila central: mezclas. Fila inferior: separación a partir de las mezclas