

Sistema de Clasificación Bayesiano basado en Múltiples Clases

Omar D. Castrillón.

Ingeniería Industrial, Universidad Nacional de Colombia
Manizales, Caldas 0017, Colombia

Jaime Alberto Giraldo

Ingeniería Industrial, Universidad Nacional de Colombia
Manizales, Caldas 0017, Colombia

William Ariel Sarache C

Ingeniería Industrial, Universidad Nacional de Colombia
Manizales, Caldas 0017, Colombia

RESUMEN

En el presente trabajo se diseña en Matlab, una metodología que permite construir un clasificador bayesiano de n clases. Este trabajo se desarrolla en dos etapas; en la primera, se establecen todos los procesos necesarios para la construcción del clasificador y entrenamiento del sistema. En la segunda etapa, se realizan las pruebas de validación cruzada, necesarias para determinar la efectividad del mismo.

Si bien, como resultado de este trabajo se obtiene un clasificador estable, con un índice de efectividad superior al 94%; es importante desarrollar el programa en un lenguaje de programación menos pesado, como C, con el fin de obtener un sistema más versátil y rápido.

Palabras claves: Error de entrenamiento, validación cruzada, media, covarianza, función discriminante, bayesiano.

1. INTRODUCCION

El objetivo de un sistema Bayesiano es saber cual es la hipótesis más probable entre varios conjuntos de datos. Si $P(D)$ es la probabilidad a priori de los datos, $P(D|h)$ su probabilidad dada una hipótesis h y se desea estimar $P(h|D)$, la probabilidad posterior de h dados los datos. Se puede plantear el siguiente teorema:

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \quad \text{Ec. (1)}$$

La hipótesis más probable o MAP (maximun a posteriori hipótesis):

$$h_{MAP} = \arg \max_{h \in H} (P(h|D)) \quad \text{Ec. (2)}$$

$$h_{MAP} = \arg \max_{h \in H} \left(\frac{P(D|h)P(h)}{P(D)} \right) \quad \text{Ec. (3)}$$

$$h_{MAP} = \arg \max_{h \in H} (P(D|h)P(h)) \quad \text{Ec. (4)}$$

Un algoritmo Bayesiano puede ser fácilmente implantado si se calculan todas las posibles hipótesis en la ecuación 1 y se selecciona la hipótesis de mayor probabilidad. En general si tiene un sistema de aprendizaje de lo general a lo específico (o al revés) que busca especializaciones más generales (o generalizaciones más específicas) se puede caracterizar

asumiendo que las hipótesis más generales (o específicas) son más probables que otras [1].

El principio básico, ilustrado en el párrafo anterior, puede ser empleado para determinar de una forma a priori la clase en la cual puede ser clasificado un dato, según las probabilidades generadas por una serie de funciones de clasificación bayesiana, previamente definidas.

Tomando como referencia estos principios, ciertos autores [2,3], han establecido funciones de evaluación Bayesiana, las cuales se basan en definir una función de probabilidad para cada una de las clases, que el algoritmo considere.

De otro lado, es importante señalar que los algoritmos basados en un enfoque bayesiano han sido usados en varias y diversas áreas de la ingeniería como:

- Formulación de nuevos métodos para la extracción y agrupamiento de características en la visión artificial; encontrándose una mayor eficiencia, robustez y flexibilidad frente a otros métodos tradicionales [4].
- Clasificación automática de patrones, presentes en los carbones colombianos.

En esta investigación se compara el modelo bayesiano, frente a otros modelos de clasificación como las máquinas de soporte vectorial y redes neuronales; encontrándose que el modelo bayesiano genera la mayor capacidad predictiva, principalmente por su robustez frente al ruido [5].

- Extracción y clasificación de posturas labiales en niños entre 5 y 10 años en la ciudad de manizales.[6]
- Identificación de voces normales y disfuncionales, en la zona centro de Colombia, a partir de su análisis acústico. Considerando, que no existe un único patrón de voz normal, en esta investigación, el clasificador bayesiano inicia con una etapa de aprendizaje sobre la normalidad de la voz en un contexto determinado, finalmente se realizan los respectivos procesos de validación con resultados superiores al 95% [7].
- Aproximación de funciones de valores discretos, capaz de expresar hipótesis disyuntivas y robustas frente a factores de ruido [8].
- Clasificación estadística, aplicada al reconocimiento de imágenes [9].
- Identificación de fallos, basados en series de reconocimiento de patrones, generados por los valores históricos de las variables observadas [10].

En general, los clasificadores estadísticos, desde un enfoque bayesiano, han sido ampliamente usados en la solución de diversas clases de problemas, principalmente por presentar una

mayor robustez frente a factores no deseados como el ruido; aspecto que los convierte en un tema de gran interés para la investigación.

2. METODOLOGIA

Sea x_i , la matriz de hiperpuntos que contiene cada una de las n clase, de tamaño N_c (muestras por clase) x D (Número de características) x C (Clases). Se procede del siguiente modo:

- Se calcula el vector de medias μ_i de x_i .
- Se calcula la matriz de covarianza \sum_i de x_i
- Se calculan los coeficientes de las funciones discriminantes para cada una de las n clases[2]:

$$W_i = \frac{1}{2} \sum_i^{-1} \quad \text{Ec. (5)}$$

$$w_i = \sum_i^{-1} \mu_i \quad \text{Ec. (6)}$$

$$w_{io} = \frac{1}{2} \mu_i^t \sum_i^{-1} \mu_i - \frac{1}{2} \ln \left| \sum_i \right| + \ln P(\omega_i) \quad \text{Ec. (7)}$$

Los coeficientes de las funciones discriminantes, son calculados para cada una de las n clases, con una base de datos de entrenamiento, previamente definidas.

- Se definen las funciones discriminantes para cada una de las n clases:

$$g_i = x^t W_i x + w_i^t x + \omega_{io}, \quad \text{Ec. (8)}$$

- Los valores de características seleccionadas y pertenecientes a la base de datos de validación son evaluados sobre cada una de las n funciones discriminantes de probabilidad generadas (Ecuación 8). Se supone que la muestra evaluada, pertenece a aquella clase cuya función de probabilidad genera el máximo valor.
- Los resultados obtenidos en el numeral anterior son comparados con una etiqueta previamente establecida, la cual indica el valor real de la clase, a la que pertenece la muestra. Como resultado de la anterior comparación se define el porcentaje de aciertos y errores que el sistema proporciona en la identificación, de las clases a las cuales pertenece la muestra.

3. MARCO EXPERIMENTAL

La experimentación de esta metodología se llevo, sobre características acústicas obtenidas sobre voces de niños normales e hipernasales, hiponasales, mixtas y otras como se especifica a continuación:

- Construcción de base de datos.** Corresponde a una muestra, que evalúa 90 niños de las 5 diferentes clases (18 pacientes por cada clase), valorados por el especialista. Cada grabación está conformada por las cinco vocales del idioma español. Las señales fueron tomadas en condiciones de bajo nivel de ruido ambiental.
- Selección del espacio Inicial de Características.** El conjunto total de características acústicas de voz consideradas son: Pitch (F_0), Jitter, Jitter porcentual (J_p), Coeficiente de Perturbación de Tono, NEP, GNE, HNR, Energía, Cruces por cero, Cepstrales, LPC y MFCC. Dada

una característica ξ , a la cual le corresponde el vector x formado por las observaciones $\{x_i; i=1, \dots, n\}$ de medida para cada una de las $k=n$ clases, se realiza la estimación de los momentos en calidad de valores de descripción de la estructura de aleatoriedad de las distribuciones empíricas:

- Parámetros de posición.** El momento inicial de primer orden (valor medio $m_{1\xi}$) se estima a efectos de ser removido de la observación, esto es, $x - m_{1\xi}$, en la medida en que se considere que este no tiene información y pueda generar un sesgo inapropiado en la interpretación de los resultados.
- Parámetros de escala.** Se consideran los siguientes: la varianza, el valor cuadrático medio, el coeficiente de variación $m_{2\xi} / \sigma_{\xi}^2$, el valor pico a pico y la desviación de mediana absoluta $\text{med} |x_i - m_{1\xi}|$. Además, se consideran los momentos centralizados de orden $l=3, \dots, 6$.
- Parámetros de forma.** Corresponden a los coeficientes de asimetría (oblicuidad) y exceso (curtosis). Además de los anteriores momentos, se pueden considerar los cumulantes de orden $m=1, \dots, 4$

En total se generan 128 variables de representación por cada palabra. Los conjuntos o subconjuntos de estas variables constituirán las entradas del clasificador bayesiano.

- Preproceso de datos.** Tiene como objetivo la disminución de la influencia y en lo posible eliminación de los errores de medida ocasionados, entre otras razones, por errores sistemáticos en el registro de las señales, fallas ocasionales en los dispositivos de medida, etc., así como el control de la homogeneidad de las propiedades estadísticas de las diferentes características del fenómeno aleatorio en análisis. El preproceso de los datos consiste del análisis de registros anómalos de cada característica y la verificación de normalidad de cada una de ellas.
- Clasificación.** Este proceso se realiza por medio de la metodología propuesta en el apartado anterior. Con el fin analizar la variación de los parámetros de clasificación y su capacidad de generalización. Se realiza una validación cruzada sobre el clasificador.
- Software.** tomando como referencia algunos autores [11]. se diseñaron los algoritmos con los cuales se programó la anterior metodología para este marco experimental.

Function error=clasibay(datos,s,tam)

```
load(datos) % Se cargan los datos de disco;
normales=normales(:,[eval(s)]);
anormales=anormales(:,[eval(s)]); %Estan todas las clases (4)
matriz=[normales(1:tam,:);anormales(1:tam,:)];
tam=size(matriz); % Matriz de muestras
n=n; %Número de clases
```

% Primer entrenamiento del sistema

```
[a(:,1),b(:,1),c(:,1)] = bayesiano(normales(1:end/2,:),1/n);
[a(:,2),b(:,2),c(:,2)] = bayesiano(anormales(1:end/8,:),1/n);
[a(:,3),b(:,3),c(:,3)] = bayesiano(anormales(end/4:end*3/8,:),1/n);
```

```
[a(:,4),b(:,4),c(:,4)]
bayesiano(anormales(end*1/2:end*5/8,:),1/n);
[a(:,5),b(:,5),c(:,5)]
bayesiano(anormales(end*6/8:end*7/8,:),1/n);
%Para mas clases se amplia la matriz
```

```
%Primera validación del sistema
```

```
matriz_v=[normales(end/2+1:end,:);
anormales(end*8/8:end*1/4,:), anormales(end*3/8:end*1/2,:),
anormales(end*5/8:end*6/8,:), anormales(end*7/8:end,:),];
%Para mas clases se amplia la matriz
target=[tar(tam/2+1:tam);tar(tam+tam/2+1:2*tam)];
```

```
matriz_v=[matriz_v target];
matriz_v=shuffle(matriz_v);
```

```
%Error del sistema
```

```
[error,clase]=errorent(a,b,c,matriz_v(:,1:end-1),matriz_v(:,end));
```

```
%Segundo Entrenamiento
```

```
[a(:,1),b(:,1),c(:,1)] = bayesiano(normales(end/2+1:end,:),1/n);
[a(:,2),b(:,2),c(:,2)] = bayesiano(anormales(1:end/8,:),1/n);
[a(:,3),b(:,3),c(:,3)] = bayesiano(anormales(end/4:end*3/8,:),1/n);
[a(:,4),b(:,4),c(:,4)] = bayesiano(anormales(end*1/2:end*5/8,:),1/n);
[a(:,5),b(:,5),c(:,5)] = bayesiano(anormales(end*6/8:end*7/8,:),1/n);
%Para mas clases se amplia la matriz
```

```
%Segunda validación del sistema
```

```
matriz_v=[normales(1:end/2,:); anormales(end*8/8:end*1/4,:),
anormales(end*3/8:end*1/2,:), anormales(end*5/8:end*6/8,:),
anormales(end*7/8:end,:),]; %Para mas clases se amplia la matriz
```

```
target=[tar(1:tam/2);tar(tam+1:tam+tam/2)];
```

```
matriz_v=[matriz_v target];
matriz_v=shuffle(matriz_v);
```

```
%Error del sistema
```

```
[error,clase]=errorent(a,b,c,matriz_v(:,1:end-1),matriz_v(:,end));
```

```
function [a,b,c]=bayesiano(clases,prob)
```

```
t=size(clases);
u=mean(clases);
r=zeros(t(2),t(2));
```

```
for(i=1:t(1))
r=r+(clases(i,:)-u)*(clases(i,:)-u);
end
```

```
%Cálculo de coeficientes discriminates
```

```
ci=r/t(1);
```

```
a=-1/n*inv(ci);
b=u*inv(ci);
c=-1/n*(u*inv(ci)*u'+log(prob)-1/n*log(det(ci)));
```

```
function [error,clase]=errorent(a,b,c,base,tar)
```

```
errorb = 0;
tam=size(base);
N=n;
tam = size(base);
for i=1:tam(1)
p = base(i,:);
for j=1:N
r(j)=evaluabayes(a(:,j),b(:,j),c(:,j),p);
end
[h,clase(i)] = max(r);
if clase(i)==0
tartrain=1;
else
if clase(i)==1
tartrain=2;
else
if clase(i)==2
tartrain=3;
else
if clase(i)==3
tartrain=4;
if clase(i)==4
tartrain=5;
else
tartrain=-1;
end
if tartrain~=tar(i)
errorb = errorb +1;
end
end
error = (errorb/tam(1))*100;
```

```
function valor=evaluabayes(a,b,c,punto)
```

```
valor=punto*(a*punto)+(b*punto)+c;
```

4. RESULTADOS Y DISCUSIONES

A continuación se presentan los resultados para cada una de las etapas propuestas en el marco experimental:

- Construcción de base de datos:** Como resultado de este proceso se generó una base de datos de 128 características, medidas en 90 niños, (18 por cada clase). Estas características se establecieron sobre cinco vocales. Las características analizadas fueron: 14 relacionadas con la frecuencia fundamental, 3 relacionadas con el Jitter, el NEP, 7 relacionadas con el GNE, 14 relacionadas con el HNR, 27 relacionadas con los coeficientes Ceptrales, 18 relacionadas con los coeficientes Mel Ceptrales, 15 relacionadas con la energía, 4 relacionadas con los cruces por cero, 15 relacionadas con la predicción lineal y 10 con los coeficientes de predicción lineal. Para un total de 128.
- Selección del espacio inicial de Características.** Este proceso generó una base de datos debidamente normalizada, con media cero y desviación estándar 1.
- Preproceso de datos** Se realizó un procedimiento de detección de valores anómalos por cada una de las características ξ . Finalizado este proceso no se encontró valores anómalos. La verificación de normalidad, el juicio

sobre la estructura gaussiana de los datos, se pudo realizar a partir de la respectiva prueba de hipótesis. En el trabajo se emplea la prueba de Kolmogorov-Smirnov.

- d. **Clasificación:** Este proceso se realizó mediante la metodología ilustrada en el punto 2. Los resultados de esta fase muestran una efectividad promedio del clasificador del 94% en la identificación de las n clases analizadas.

Si bien, otros clasificadores basados en redes neuronales, máquinas de soporte vectorial, y otros, parecen obtener resultados ligeramente superiores, en condiciones de normalidad, es importante resaltar que esta metodología es más robusta ante factores de ruido; además puede ser empleada para reducir parámetros acústicos y seleccionar los más efectivos [12,13,14,15].

5. CONCLUSION

En este clasificador se obtiene una efectividad del 94%, aspecto que lo convierte en un algoritmo confiable en el reconocimiento de una clase a priori, a la cual puede pertenecer una característica. Sin embargo y aunque los algoritmos diseñados tienen una complejidad lineal, se hace necesario migrar este software a un lenguaje más versátil como C, ó Delphi. El cual permita una mayor versatilidad y rapidez del sistema. Este aspecto permitirá definir una matriz de hiperpuntos más amplia la cual permita analizar un mayor número de clases en sus fases de entrenamiento y validación.

6. AGRADECIMIENTOS

Se agradece la colaboración prestada a todos los miembros del Grupo en Control y procesamiento digital de señales de la Universidad Nacional de Colombia sede Manizales.

7. REFERENCIAS

- [1] E. Morales, **Descubrimiento de conocimiento en bases de datos**, Monterrey: Instituto Tecnológico de Monterrey. Pub., 2001.
- [2] R. O. Duda, **Pattern Classification**, New York: Wiley Interscience. Pub., 2002.
- [3] S. Theodoridis, **Pattern Recognition**, Oxford: Elsevier. Pub., 2005.
- [4] Q. Cazorla. Miguel, **Un enfoque bayesiano para la extracción de características y agrupamiento en visión artificial**, Alicante: Universidad de Alicante. Pub., 2000.
- [5] C.J. López, “Comparación de modelos de clasificación automática de patrones texturales de minerales presentes en los carbones colombianos”, **Dyna**, Vol. 72, No. 146, 2005, pp 115 -124.
- [6] J. A. Salazar, **Extracción y clasificación de posturas labiales en niños entre 5 y 10 años de la ciudad de manizales para el control del diagnostico de la población infantil con labio y/o paladar hendido corregido**, Manizales: Universidad Nacional de Colombia. Pub., 2004.
- [7] O. D. Castrillon, **Método para la discriminación entre voces normales y disfuncionales basado en la selección efectiva de parámetros acústicos de voz. Aplicación en la zona centro de Colombia. Tesis Doctoral**, Valencia: Universidad Politécnica de Valencia. Pub., 2004.

- [8] A. H. Guerra, **Aprendizaje automático. Árboles de decision**, Veracruz: Universidad veracruzana. Pub., 2004.
- [9] L. M. Bergasa, **Clasificadores estadísticos aplicados al reconocimiento de imagines**, Alcalá: Universidad de Alcalá. Pub., 2.004.
- [10] C.J. Alonso, **Técnicas de aprendizaje para la identificación de fallos en sistemas dinámicos**, Valladolid: Grupo de sistemas inteligentes. Pub., 2004.
- [11] O. D. Castrillón, W.A. Sarache et all, “Algoritmo de Clasificación Bayesiano”, **6ta Conferencia Iberoamericana en sistemas, Cibernética e Informática**, Vol. 1, 2007, pp. 137 – 140.
- [12] O.D. Castrillon, “Selección de características acústicas en la identificación de hipernasalidad por medio de algoritmos evolutivos”, **Congreso Anual de Ingeniería Biomédica**, Vol. 1, 2006, pp. 185-188.
- [13] C.M. Giorgio, M. Zaffalon, “JNCC2: An extension of naive Bayes classifier suited for small and incomplete data sets”, **Environmental Modelling & Software**, Vol. 23, No. 1, 2008, pp. 960 –961.
- [14] J. Chen, H. Huang, et All, “A Selective Bayes Classifier for Classifying Incomplete Data Based on Gain Ratio”, **Knowledge-Based Systems**, Vol. 21, No. 7, 2008, pp. 530-534.
- [15] J.H. Hong, J.K. Min, U.G. Cho, Sung-Bae Cho, “Fingerprint classification using one-vs-all support vector machines dynamically ordered with naïve Bayes classifiers”, **Pattern Recognition**, Vol. 41, No. 4, 2008, pp. 662 – 671.