

Comparación de clasificadores de género: Redes Neuronales MLP y Teoría de la Información –ITL–

Comparison of gender classifiers: MLP Neural Networks and Information Theory –ITL–

Ing. Efredy Delgado Aguilera M.Sc.¹

¹ *Estudiante Doctorado Ingeniería Eléctrica, Universidad de Chile.*

Fecha de recepción del artículo: 03/08/2010; Fecha de aceptación del artículo: 25/01/2011

Resumen

En el presente trabajo se comparan dos métodos para la clasificación de datos concretamente en el caso de clasificación de género por medio de imágenes de rostros humanos, uno por redes neuronales multicapa y otro mediante teoría de la información, usando información mutua cuadrática de Cauchy.

Se muestran resultados del uso de redes neuronales y su configuración así como en el caso de información mutua y los casos de acierto en sus fases.

Palabras clave

Teoría de la Información, Redes Neuronales, Clasificación de Género, Información Mutua, Información Mutua Cuadrática de Cauchy.

Abstract

This paper compares two methods for data classification, specifically for gender classification using images of human faces, the first method for multilayer neural networks and other means of information theoretic learning, using Cauchy's quadratic mutual information. This paper shows results of using neural networks and their settings as well as in the case of mutual information and cases of success in its early stages.

Keywords

Theoretic Information Learning, Neural Networks, gender classification, Mutual Information, Cauchy's quadratic mutual information Universidad de Chile Comparación de Clasificadores de Género - Redes Neuronales vs ITL.

Introducción

El reconocimiento de rostros mediante fotografías o secuencias de imágenes, es un problema inmanente a la visión computacional, algunas metodologías reúnen estrategias como minería de datos y reconocimiento de patrones, lo que permite exigir de quienes lo ejecutan, un panorama amplio en la aplicación de diversas técnicas.

El planteamiento al abordar este tema es tener bases sobre manejo de imágenes, con su respectivo manejo de bases de datos que permitan entrenar algún sistema de aprendizaje que luego va a ser aplicado en la solución al problema de clasificación de rostros.

La presentación se realiza en torno a tomar redes neuronales como sistema de aprendizaje y aplicar ciertas preguntas dentro del contexto de las imágenes, ¿de quién es la imagen? Explorar el conjunto de

bases de imágenes y determinar la identificación de la imagen, asignándole características únicas, ¿Es la imagen la buscada? Comparación formal entre cada una de las imágenes, o a partir de una lista de datos determina si la imagen buscada se encuentra en el universo establecido.

¿Para qué reconocer rostros?

En vigilancia y seguridad: Identificar a un individuo sospechoso que pueda aparecer en público o en lugares neurálgicos para la seguridad. Reconocimiento de intrusos en el ámbito de la vigilancia doméstica.

En las interfaces hombre-máquina: Para restringir accesos a centros de cómputo, para unidades financieras. Bloquear posibles suplantaciones a la hora de acceder a edificios privados. En la seguridad vial, la aplicación de detectores de expresiones faciales, interfaces capaces de la detección de cansancio en conductores.

Preservación de datos de medios visuales así como recuperación de historias filmicas, y condensación de información de diversa índole.

¿Qué es una red neuronal?

Para entender las redes neuronales artificiales (RNA), se debe establecer cómo funciona una red neuronal biológica, porque la estructura del cerebro se fundamenta en la composición de millones de células interconectadas entre sí mediante la sinapsis, la unidad fundamental con la cual el cerebro percibe los estímulos físicos y químicos de la naturaleza.

La unidad básica del sistema nervioso son las neuronas, las neuronas tienen tres componentes principales: las dendritas, el cuerpo y el axón. Las dendritas son los elementos receptores de la red, encargados de llevar las señales eléctricas de las excitaciones fisicoquímicas. El cuerpo conduce y suma todas las señales eléctricas desde la entrada. El axón es una fibra larga que permite la conexión desde el cuerpo hasta otras neuronas, se ramifica en su extremo en pequeños bulbos finales que

casi tocan las dendritas de las células vecinas. Un esquema para describir una red neuronal biológica es la que se presenta en la Figura 1.

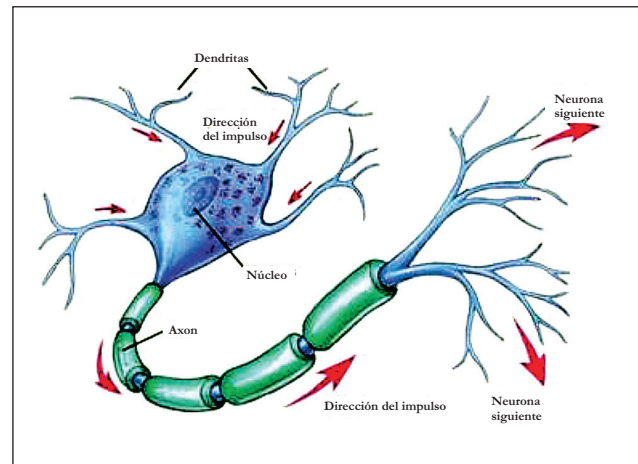


Figura 1. Componentes de una Red neuronal biológica (tomado de <http://blog.pucp.edu.pe/>).

Las redes neuronales funcionan de igual forma como lo hace el resto del cuerpo humano: mediante impulsos eléctricos y reacciones químicas. Los impulsos eléctricos que usan las redes neuronales para intercambiar información con las demás, viajan por el axón que hace contacto con las dendritas de la neurona vecina mediante las sinapsis. La intensidad de la corriente transmitida depende de la eficiencia de la transmisión sináptica.

El impulso eléctrico es transmitido mediante las dendritas y conducido por el axón hacia otras conexiones entre otras neuronas. De manera que el estímulo percibido por las dendritas sea la causa su reacción, en los órganos o tejidos vinculados a esas neuronas, depende de la intensidad del estímulo y de la ponderación del mismo.

Una red neuronal artificial es un modelo que emula el comportamiento de las redes neuronales biológicas, describiéndolas como una unidad de proceso que actúa en paralelo, que consta de unas entradas x_i equivalentes a las dendritas en su paralelo biológicos, dichas entradas están ponderadas por pesos w_i (ver Figura 3), las cuales responden a la

forma como los impulsos eléctricos son evaluados y se combinan con la función de red, la cual define el nivel de potencial de la neurona. La salida de la función de la red es evaluada en la función de activación que da lugar a la salida de la unidad de proceso.

En forma simplificada el esquema de la red neuronal artificial corresponde a la mostrada en la Figura 2.

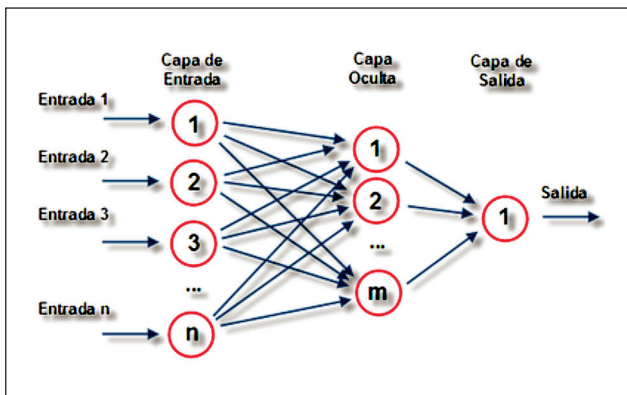


Figura 2. Esquema de una red neuronal artificial (tomado de <http://www.ikkaro.com/redes-neuronales>).

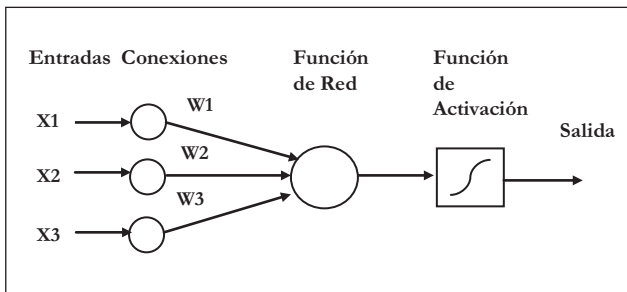


Figura 3. Unidad de proceso típica.

La manera por la cual las redes neuronales artificiales realizan el proceso de cambio de la ponderación de las entradas mediante el cambio de pesos se denomina **aprendizaje de la red**, el modo de aprendizaje se dictamina de acuerdo a las condiciones de la misma red o a su configuración, entendiéndose esto último como se organizan las neuronas con otras unidades, estableciendo la arquitectura de la red neuronal, dentro de los tipos

de entrenamiento más importantes se encuentran las estrategias de retropropagación (*back-propagation*), la regla delta (*delta rule*) y el aprendizaje de Kohonen por citar algunos. Existen tantas formas de aprendizaje como variaciones en la arquitectura de las redes; las reglas de aprendizaje se subdividen en dos categorías: el aprendizaje supervisado y el no supervisado, el principio del aprendizaje supervisado permite establecer la relación entre la salida deseada ante una entrada determinada, mientras el aprendizaje no supervisado en el cual la salida es evaluada sin una guía relacionada con la entrada.

Modelos de una red neuronal simple

Los modelos de las redes neuronales pueden ser de tipo lineal, en donde la función de red, es una combinación lineal de los atributos de entrada y los pesos:

$$(1) \quad y = w_0 + w_1x_1 \dots + w_nx_n$$

La función de activación con la que la neurona responderá ante un estímulo, puede darse de diferentes formas: lineal, escalón, tangencial o hiperbólica.

Introducción a la Teoría de la Información ITL (*Information Theoretic Learning*)

De acuerdo a la definición de José Principe “El aprendizaje por teoría de la información es un paradigma de entrenamiento que usa teoría de la información, muestreo, (entropía) para adaptar redes con parámetro lineales como no lineales. La teoría de la información ha sido aplicado en optimización, pero requiere de suposiciones a priori, de funciones de densidad de probabilidad (P.D.F). Sin embargo, ITL no es paramétrico, i.e. se explota la información contenida del muestreo”².

La teoría de la información surge a principios de la década de los 40's, cuando se buscaba hacer más eficientes las formas de comunicación, sus canales y la medición de su capacidad. La teoría de la información usa teoría de probabilidad, para realizar

las mediciones de información y contenido, cuanto más imprevisto es el suceso analizado, más cantidad de información puede originar, el contenido de la información de un suceso es igual a la función decreciente de la probabilidad de su aparición.

Para convertir las probabilidades en información Shannon usó el logaritmo de la inversa de la probabilidad. De acuerdo con la teoría expuesta por Shannon, si se tienen un conjunto de sucesos n , uno de los cuales va a presentarse con certeza, por lo que cada uno tiene una probabilidad de que ocurra:

$$(2) \quad \sum_{i=1}^n x_i = 1 \quad \sum_{i=1}^n x_i = 1$$

El contenido de la información definido por Shannon es:

$$(3) \quad H(x) = \sum_{i=1}^n x_i \log \frac{1}{x_i} H(x) = \sum_{i=1}^n x_i \log \frac{1}{x_i}$$

Metodología

La clasificación de rostros femeninos como masculinos arranca con la disposición de una base de datos que permita la extracción de características geométricas de diversa índole, para el caso del autor citado, define 18 características geométricas [1].

En primera instancia, se establece una base de datos de prueba del Georgia Tech, con imágenes de mujeres y hombres, la cual posee 50 imágenes de personas en formato JPEG, cada individuo con 15 imágenes a color, con variaciones de iluminación, expresiones y escales diferentes y orientaciones.

Se definen marcas para establecer unas características en relación a mediciones geométricas de los rasgos de rostro.

A partir de los datos anteriores se estructura una base de datos, de (n) patrones con valores escalares para todos los (m) sujetos, junto a su clasificación de género.

Una vez se obtiene la base de datos, se diseña y entrena la red neuronal, y se establece el desempeño,

se diferencian las características de mayor relevancia en la determinación de género y se comparan los resultados con otros métodos y con resultados de artículos relacionados.

Metodología MLP (*Multilayer Perceptron*)

Para la selección de las características relevantes ingresadas a la red neuronal, se escoge una red con arquitectura de retro propagación del error que permita rescatar la información de los pesos asignados a cada entrada. La red anterior, corresponde a una con N entradas ($N=13$ en este caso) sin capa oculta y una salida, con función de activación lineal.

Lo anterior permite conocer la ponderación de cada peso en la clasificación. Para poder determinar los coeficientes (pesos asignados a las entradas), se entrena la red y se almacena la que presente el mejor desempeño. Una vez obtenido los pesos, se ordenan las entradas de acuerdo a sus pesos en orden decreciente. Se escogerán las características que presenten pesos mayores que un umbral $q = 0.5$. Las demás se descartan.

Una vez que se tengan las características más relevantes se entrenan nuevas la misma topología y distintos parámetros de aprendizaje, comparándose el porcentaje de clasificación correcto con el obtenido con todas las entradas. Luego se varía la arquitectura de la red y se entrenan con las características seleccionadas y, separadamente, con todas las características, para comparar los desempeños. Finalmente, se establece la pérdida de información al reducir el número de características como el porcentaje de variación promedio entre las dos redes clasificadoras (la que recibe todas las entradas y el espacio reducido).

Algoritmo utilizado: ITL

En cuanto al algoritmo utilizado, el primer paso fue la extracción de las distancias geométricas del rostro de la base de datos, construcción de los funcionales de relevancia y de redundancia en el simulador, incorporados estos dos funcionales son desarrollados para determinar la información

mutua cuadrática de Cauchy, se emplea un sistema de selección múltiple de comparación para establecer la información mutua entre cada uno de las características a evaluar para conformar de la totalidad de características, un conjunto reducido que permita mantener la exactitud con la cual se clasifica el género, a partir de este nuevo conjunto se introducen al clasificador para que determine el proceso de clasificación y poder hacer la comparación a partir de las características enteras.

Las funcionales que tuvieron que ser descritas para el simulador están:

$$(4) \quad I_{CS}(X; C) = \log\left(\frac{V_{(C,X)}^2 V_{(C)}^2 V_{(X)}^2}{V_{(C,X)}^2}\right)$$

La información mutua de Cauchy que permite definir la distancia entre las probabilidades marginales asociadas a cada una de las clases y la probabilidad conjunta de las mismas, esta distancia se basa en la distancia de Cauchy

$$(5) \quad |\langle x, y \rangle|^2 \leq \langle x, x \rangle \cdot \langle y, y \rangle \quad |\langle x, y \rangle|^2 \leq \langle x, x \rangle \cdot \langle y, y \rangle$$

El cual establece que un par de vectores, de punto interno real, equivale a tomar la raíz cuadrada de la desigualdad en ambos lados, lo que lleva a establecer la norma de los vectores, escribiendo la ecuación como:

$$(6) \quad |\langle x, y \rangle| \leq \|x\| \cdot \|y\| \quad |\langle x, y \rangle| \leq \|x\| \cdot \|y\|$$

La siguiente función a desarrollar es la que tiene que ver con la relevancia de características (Features Relevance Criterion FRC).

De igual manera, con el criterio de redundancia de las características (FSC) se mide la similitud del subconjunto de características S y una característica f_m ; de tal manera, que si la característica f_m es muy similar a f_i que además pertenece a al subconjunto S , aquélla debe ser similar (o redundante).

$$(7) \quad FRC(f_m) = I(S + f_m; C)$$

$$(8) \quad FSC(f_m) = \arg \max_{f_i \in S} \left(\frac{I(f_m; f_i)}{H(f_i)} \right)$$

La redundancia de una característica debe ser igual o menor a 1 sí y solo sí la información esta característica no contiene más información que la dada por S . Cuando la redundancia es suficientemente grande tal que $FSC \geq \theta$, umbral determinado por los documentos revisados se ha tomado con un valor de 0.9, y por lo tanto de acatarse este umbral, la característica en mención puede tomarse como redundante y por lo tanto no se adiciona al subconjunto S .

El proceso de selección de características se basa en el siguiente diagrama de flujo, para el cual permite definir del conjunto total de características un método con el cual elegir las más relevantes mediante el umbral θ , descartando las redundantes. El proceso empieza con un conjunto S que es vacío y alojará el subconjunto de características, se selecciona la característica del conjunto total F con base en el criterio de relevancia, ésta se elimina del conjunto de origen F y se evalúa con respecto al criterio de redundancia de sobrepasar esta evaluación, la característica se ubica en el subconjunto S ; por último el criterio de parada se define como un número apropiado de características seleccionado identificado como B en el cual, debe satisfacer que este índice sea igual o mayor a un escalar por el máximo Criterio de Relevancia de Característica (FRC). El diagrama de flujo del procedimiento se presenta en la Figura 4 (pág. 92).

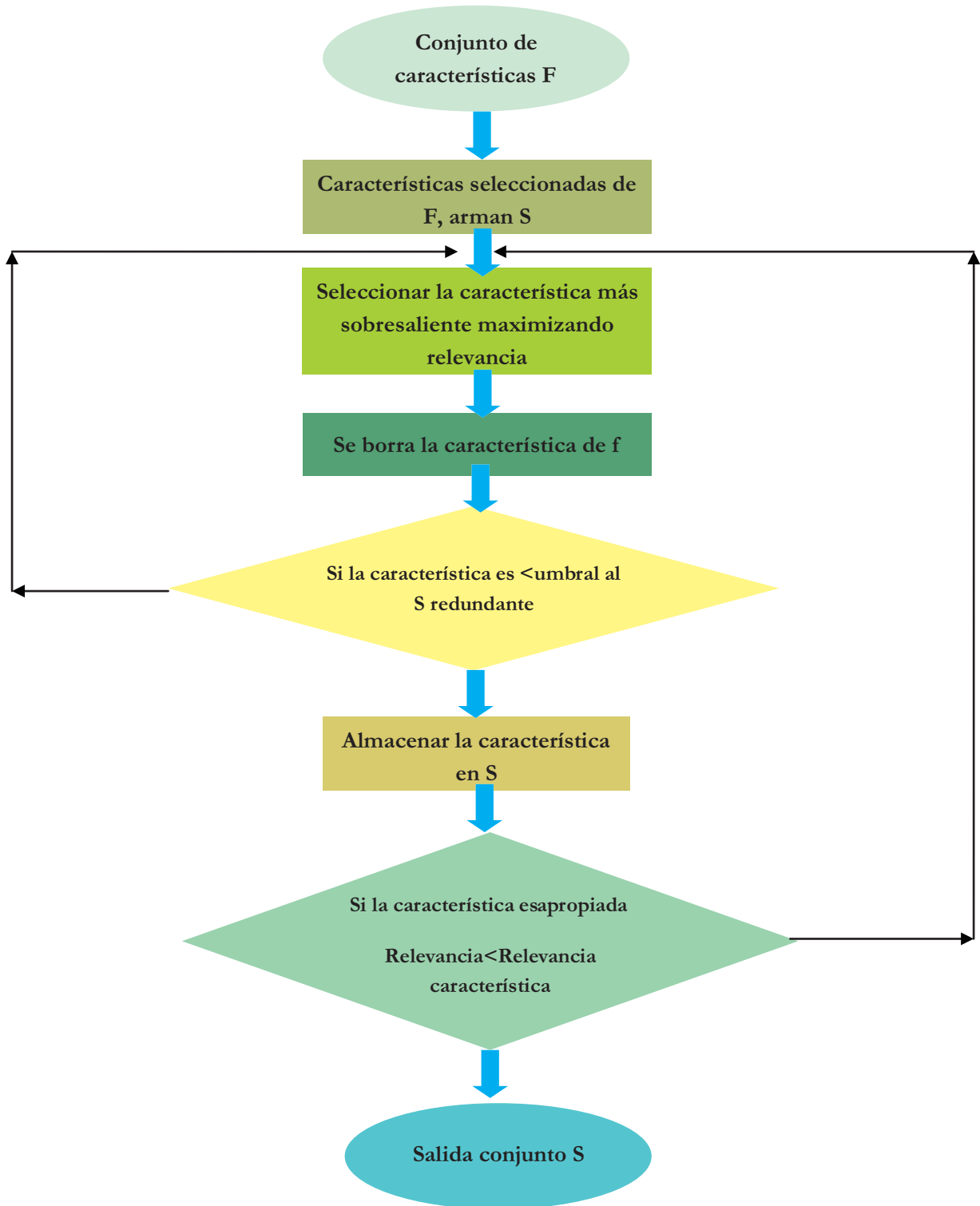
Bases de datos

Se tienen 236 imágenes extraídas de las bases de datos FERET y Yale, además de unas seleccionadas de Internet (las que cumplen con iluminación suficiente y vista total del rostro), de las cuales 106 son mujeres, conjunto que se subdividió en 142 para el conjunto de entrenamiento (55 mujeres), 47 para el de validación (24 mujeres), y 47 para el de prueba (27 mujeres).

La demarcación realizada manualmente, por lo que la base de datos es reducida y por ello no se uso la cantidad disponible de las bases como FERET.

Las características son:

Figura 4. Diagrama de flujo del método establecido para el uso de criterios de relevancia y redundancia (gráfica realizada por el autor).



1. Razón alto ancho de nariz.
2. Razón entre el producto distancia entre ojos ancho nariz y el ancho de la cara.
3. Distancia promedio entre extremo de cejas y borde de ojos.
4. Razón entre la suma de altura de cejas y suma de ancho de cejas.
5. Razón entre la suma de altura de ojos y suma de ancho de ojos.
6. Ancho mandíbula.
7. Alto nariz.
8. Altura promedio de ojos.
9. Altura promedio cejas.
10. Razón entre media de la región mentón y promedio de las demás tres regiones.
11. Promedio de las varianzas de las cuatro regiones.
12. Promedio de las desviaciones estándar de las cuatro regiones.
13. Razón entre el ancho de la nariz y el ancho de la mandíbula.

Se debe recordar que todas las distancias están normalizadas a distancias característica de cada rostro, es decir, las distancias verticales se normalizan con el alto de la cara (distancia entre extremo superior de la frente y extremo inferior del mentón), y las distancias horizontales se normalizan con el ancho de la cara, que se mide como la distancia entre los extremos de los maxilares (para una vista frontal, puede ser la distancia entre orejas).

Definición de las salidas deseadas

Las salidas deseadas que el clasificador va a definir son binarias las cuales se etiquetan como “0” hombre, “1” mujer.

Desarrollo

Dentro de los recursos a utilizar están la base de datos de Georgia Tech, modificada debido a la escasez de imágenes con rostros femeninos, se acudió entonces a imágenes similares que contengan rostros femeninos, utilizando un conjunto total de 87 imágenes de las cuales 47 son imágenes femeninas y 40 imágenes masculinas.

Se usan los toolbox de redes neuronales y de imágenes de Matlab.

Se desarrolló un programa que recibe las mediciones y conforma una base de datos.

Las mediciones geométricas establecidas fueron 13, se adicionó como variante un método de escalamiento por región el cual consiste en tomar tres zonas del rostro: mejillas, frente y mentón (Figura 5), a las cuales se les aplica una descomposición RGB, se hacen una ponderación a cada uno de ellos, de tal manera que se establezca una norma euclidiana, obteniendo un escalar que permita calcular las medias de tonalidades encerradas en cada zona. Las características geométricas se unen con este conjunto de parámetros, con los cuales se alimenta la red neuronal.

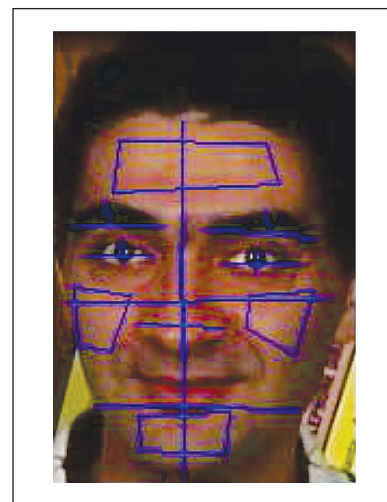


Figura 5. Regiones de tonalidades y medidas geométricas (figura tomada de Georgia Tech, modificación hecha por el autor).

La red usada para la clasificación, estaba estructurada con 13 entradas, 16 nodos en la capa oculta, y una salida, la cual va a optar por 0 ó 1, si es el caso de hombre o mujer respectivamente; esta red implementada como de retro propagación con entrenamiento supervisado y con gradiente descendente en la que se probaron diferentes coeficientes de aprendizaje, la funciones de salida es

un función lineal mientras que la función de la capa oculta es una sigmoide. En ella se incorporaron 240 ejemplos, 144 con conjunto de entrenamiento, de las cuales 55 son imágenes de mujeres y 48 para el conjunto de validación, de las cuales 27 han sido mujeres; el conjunto de prueba es de 48, con 24 imágenes de mujeres.

Resultados y análisis

Usados en la metodología de ITL:

Las características relevantes después del proceso de selección se definen de la siguiente manera, del más relevante al menos relevante:

- Característica 13: Razón entre ancho de nariz y ancho de mandíbula.
- Característica 12: El promedio entre las desviaciones estándar de las cuatro regiones faciales (rectángulos seleccionados en frente, mentón y dos mejillas).
- Característica 11: Promedio de las varianzas de las cuatro regiones faciales antes mencionadas.
- Característica 10: La razón entre la media de la región del mentón y el promedio de las medias de las otras tres regiones faciales.
- Característica 9: El promedio del alto de la ceja.
- Característica 8: El promedio de la altura de ojos.
- Característica 7: El alto de la nariz.

Hay que recordar los datos se encuentran en forma normalizada, lo que significa que las medidas verticales se dividen por el alto de la cara y las medidas horizontales se dividen por el ancho de la cara.

Con estos datos de características se logró establecer un porcentaje de acierto del 85 % con una tasa de aprendizaje del 0.001 con 16 neuronas en la capa oculta, se realizaron 10 épocas. El resultado global en tasa de aciertos de falsos positivos y falsos negativos se tiene en las Tablas 1 y 2.

Se puede apreciar que el aumento del error pasó del 4% al 14 % con un porcentaje de acierto del 85% frente al registrado con la totalidad de las características que fue del 97%, esta solución está planteada con la mejor configuración de la red encontrada.

Resultados en MLP

Haciéndose 40 simulaciones para cada red, se obtienen los siguientes resultados:

Se entrena la red de topología descrita anteriormente de parámetros $u=\{0.1;0.01\}$ y $\sigma = 0$. Obteniéndose el mejor porcentaje de clasificación para $u=0.1$ siendo de 65.83% en promedio.

Las características y sus pesos correspondientes se entregan en la Tabla 3.

Tabla 1. Resultados sin selección de características.

13 características							
	vp	Fp	vn	fn	Total	Aciertos	Error
Entrenamiento	47,1	10,9	76,1	7,9	142	86,76%	1,24%
Validación	24	13,5	10,5	0	48	71,88%	28,13%
Prueba	26	1	19	1	46	97,83%	4,35%

Tabla 2. Resultado con selección de características usando ITL.

7 características							
	vp	Fp	vn	fn	N	Aciertos	Error
Entrenamiento	41,9	13,8	73,2	13,1	142	81,06%	18,94%
Validación	21,2	3,2	20,8	2,8	48	87,50%	12,50%
Prueba	23,6	3,3	16,7	3,4	47	85,74%	14,26%

Tabla 3. Pesos de la red obtenidos para cada característica.

Regla N°	Peso
11	0,9738
9	0,8675
5	0,8408
3	0,814
10	0,7996
2	0,6091
13	0,3718
6	0,3144
7	0,2128
12	0,1681
4	0,1269
8	-0,1875
1	-0,3698

Las reglas seleccionadas en orden descendiente según el peso asociado son:

- Característica 11: Promedio de las varianzas de las cuatro regiones.
- Característica 9: Altura promedio cejas.
- Característica 5: Razón entre la suma de altura de ojos y suma de ancho de ojos.
- Característica 3: Distancia promedio entre extremo de cejas y borde de ojos.
- Característica 10: Razón entre media de la región mentón y promedio de las demás tres regiones.
- Característica 2: Razón entre el producto distancia entre ojos ancho nariz y el ancho de la cara.

Donde se ve que sí bien se reduce a la mitad la cantidad de entradas a la red, se requieren de todas las mediciones hechas para poder realizar la detección.

Al entrenar la red con la topología descrita primeramente (sin capa oculta y función de activación lineal), se obtienen en promedio los siguientes porcentajes de acierto para distintas tasas de aprendizaje ($\sigma = 0$). Las Tablas 4 y 5 representan el desempeño de los dos conjuntos de

entrada para el modelo más simple de perceptrón, n entradas, sin capa oculta, con una salida y una función de activación lineal, variando solo en las tasa de aprendizaje.

Tabla 4. Desempeño de la red con los distintos conjuntos de entrada.

U=0.1

N° Características	% Aciertos promedio
6	67,2825
13	65,8375

Tabla 5. Desempeño de la red con los distintos conjuntos de entrada.

U=0.01

N° Características	% Aciertos promedio
6	64,785
13	59,2925

Notándose una leve mejoría en la clasificación, indica que ante un clasificador lineal, las entradas filtradas, en promedio, sólo aportan ruido al sistema.

La nueva topología del clasificador que considerará los dos conjuntos de entrada por separado es N:10:1, con N entradas ($N = \{6,13\}$), 10 neuronas en la capa oculta y una en la de salida; La función de activación de la capa oculta es una sigmoideo, mientras que la de la salida es lineal.

Se obtuvieron los siguientes resultados para los dos distintos valores de tasa de aprendizaje u. Las Tablas 6 y 7 representan el desempeño de los dos conjuntos de entrada para la topología de entrada con n entradas, 10 elementos en capa oculta, con una salida y una función de activación sigmoideal, variando sólo en las tasa de aprendizaje.

Tabla 6. Desempeño de la red con los distintos conjuntos de entrada.

$$U=0.1$$

Nº características	% Aciertos promedio
6	88,08
13	88,1325

Tabla 7. Desempeño de la red con los distintos conjuntos de entrada.

$$U=0.01$$

Nº características	% Aciertos promedio
6	82,865
13	86,4325

Se puede ver que la red que tiene $N=13$ entradas presenta un desempeño mejor frente a ambas tasas de aprendizaje, siendo leve la diferencia para el caso de $u=0.1$ ($\varepsilon=88.1325-88.08=0.0525\%$).

En el algoritmo usado fundamentado en ITL, se muestran más características que analizando los pesos sobre la red MLP, sin embargo no presenta mejoras en el desempeño, el cual es aproximadamente de un 87% de aciertos.

Mejoras en el trabajo

La variante implementada es decir, la combinación de mediciones geométricas con las de tonalidad y de dos vertientes de la bibliografía revisada.

Si bien las características tomadas no han sido aún ponderadas, se cuenta con un clasificador que tiene el 97,2% de desempeño, comparable a los mencionados en documentos con similitud en conjuntos de prueba.

Un punto a mejorar es obtener las características, porque hasta el momento sólo se han obtenido con un método manual, no por un método automático. Adicionar o refinar la estimación de los patrones

utilizados en la conformación de nuevos conjuntos de datos seleccionados.

Conclusiones

La idea del trabajo es la identificación de características de mayor información en la clasificación de género. Es por ello, que la base de datos se estructuró de modo que cada imagen aporte características completas, es decir, libres de ruido.

La unión de elementos geométricos junto a zonificaciones de tono del rostro pueden ser positivas en la medida en que se explotan más características del rostro, lo que ayuda a la clasificación de género, es trivial que los tonos en el rostro femenino para el mentón suele ser homogénea, por lo que fue éste, un indicador usado para ser combinado junto a las medidas geométricas contempladas en las referencias.

Se asociaron razones de distancias entre las características más importantes, se destaca que los rasgos más importantes permiten reducir las características con un alto porcentaje de certeza, siendo ésta menor a la encontrada con la totalidad de las zonas.

El procedimiento usado elimina las debilidades que la pose, siempre y cuando no sea una imagen de perfil, en cuyo caso, no existe la posibilidad de medir razones de alturas entre cejas, y distancia entre los ejes de los ojos.

Se pretendió utilizar análisis de componentes principales, para obtener una reducción en las características de entrada; sin embargo, queríamos conservar la identificación de los datos para definir cuál era la incidencia de cada característica sobre los clasificadores. El inconveniente de PCA radica en que la reducción o proyección a componentes principales mezcla las características cosa que no que pretendía, porque se quería era identificar cual era el peso sobre la decisión del clasificador debido a cada una de las características, por lo que se optó usar funciones de información mutua determinadas en la relevancia y redundancia de las características.

Si bien se aumentaron la conformación de la base de datos, que en una primera instancia era de sólo 88 imágenes a 240 imágenes, con la homogenización de rostros masculinos y femeninos, en el conjunto de prueba el porcentaje de error creció desde 4% con las 13 características a 14% utilizando las siete entregadas por IITL.

Se utilizaron dos funcionales de maximización de relevancia de características y de minimización redundancia de características basadas en la información mutua cuadrática.

Se puede ver que el criterio IITL escogido entrega las características que aportan mayor información al proceso de discriminación entre géneros, siendo las características descartadas sólo de ajuste fino. Para poder considerar mejor un ajuste fino, con sus características importantes, se deben modificar los parámetros en el algoritmo de selección de características. Lo anterior se debe hacer de modo que se pueda obtener los porcentajes de certeza deseados.

Se aprecia también, en términos generales, que hay patrones faciales que son de mayor relevancia para la detección del género, siendo los más relevantes según el criterio utilizado los que entregan información de variaciones de color entre las zonas descritas (mentón, mejillas y frente), siendo coincidente con la existencia o no de barba; Luego los parámetros que entregan información sobre la relación de tamaño entre ancho de nariz y mentón, y la altura promedio de las cejas entre otras.

El método MLP comparte información junto a IITL en seleccionar características que poseen mayor información, entre las que se encuentran las características 9, 10, 11.

Referencias bibliográficas

1. Golomb D.T. Lawrence and T.J Sejnowski, SEXNET: A neural Network Identifies Sex from

Human Faces, in Advances in Neural information processing Systems vol3, 572-577 1990.

2. Huiyu Zhoua, Yuan Yuanb, Abdul H. Sadkaa Application of semantic features in face recognition. Pattern Recognition, 41 (2008) 3251-3256.

3. Mäkinen, E., & Raisamo, R. (2008). Evaluation of Gender Classification Methods with Automatically Detected and Aligned Faces. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 30(3), 541-548. Sitio web: <http://search.ebscohost.com>, doi:10.1109/TPAMI.2007.70800

4. Xudong Xie, Kin-Man Lam, An efficient illumination normalization method for face recognition Pattern Recognition Letters 27 (2006) 609–617.

5. Karungaru stephen, fukumi minoru and norio akamatsu, Face recognition in colour images using neural networks and genetic algorithms International Journal of Computational Intelligence and Applications Vol. 5, No. 1 (2005) 55–67

6. F. Smach, Design of a neural networks classifier for face detection, Journal of Computer Science. FindArticles.com. 04 Oct, 2009.

7. Gutta Srinivas, Wechler Harry, Phillips Jhonathon Gender and ethnic classification of face images. Third IEEE International Conference on Automatic Face and Gesture Recognition 1998 , ISBN: 0-8186-8344-9

8. Chow Tommy W. S. and D. Huang. Estimating Optimal Feature Subsets Using Efficient Estimation of High-Dimensional Mutual Information, *IEEE transactions on neural networks*, vol. 16, no. 1, january 213-225, 2005.