

Fusión de clasificadores débiles euclidianos, FDA y SVM por a posteriori confidence classification (APCC)

Euclidian, FDA and SVM weak classifiers fusion using a posteriori confidence classification (APCC)

Edwin Alberto Silva-Cruz
Ph. D (c) Ingeniería Electrónica
Universidad Industrial de Santander
Bucaramanga, Colombia
edwin.silva.c@gmail.com

Carlos Humberto Esparza-Franco
M. Sc. (c) Ingeniería Electrónica
Universidad Industrial de Santander
Bucaramanga, Colombia
carlosesfra@gmail.com

Resumen– Los sistemas de clasificación binario y multiclase presentan inconvenientes cuando existe traslape importante entre las clases, representación insuficiente de las clases o asimetría en la representación. Sistemas de clasificación sofisticados, incluyendo SVM (máquinas de soporte vectorial) y SVM-RBF (máquinas de soporte vectorial con funciones de base radial) pueden tener inconvenientes en la generalización de estos problemas, de manera que la obtención de una clasificación exitosa tiene inconvenientes importantes. En este trabajo se muestra cómo el uso de clasificadores por métricas más simples puede ayudar al sistema global mediante fusión de criterios usando el algoritmo APCC (A Posteriori Confidence Classification). El algoritmo APCC determina la confiabilidad individual de cada parámetro y cada sistema de clasificación y otorga ponderación a posteriori a cada clasificador en función de su salida. Los protocolos desarrollados fueron probados tanto en datos simulados como en datos reales de parámetros VPOEM y TPOEM (Volumetric Patterns of Oriented Edge Magnitudes y Temporal Patterns of Oriented Edge Magnitudes) para la representación de la expresión facial. En ambos casos el uso de APCC y fusión de clasificadores permitió incrementar significativamente la tasa de clasificación.

Palabras clave– Sistemas de clasificación, problemas multiclase, características débiles, clasificación débil, APCC

Abstract– The 2-class and multiclass classification systems have important issues when there is overlapping between the samples, insufficient representation of the classes or asymmetrical data representation. Sophisticated classification systems such as SVM and SVM-RBF may have generalization problems, so it is complicated to obtain successful classifiers. In this work it is shown how the use of classification fusion of simpler classifiers may improve the overall classification by using APCC (A Posteriori Confidence Classification). APCC defines the individual reliability of each parameter and each classification system per parameter, and produces a posteriori weight to each classifier according to its output. The de-

veloped protocols were tested using simulated data and real data from TPOEM (Temporal Patterns of Oriented Edge Magnitudes) and VPOEM (Volumetric Patterns of Oriented Edge Magnitudes) for facial expression representation. In both cases the use of APCC and classifier fusion allowed to improve the classification accuracy.

Keywords– Classification systems, multiclass problems, weak features, weak classification, APCC.

1. INTRODUCCIÓN

En sistemas de clasificación, el desbalance en los datos y el traslape entre las clases de datos usados en entrenamiento son dos problemas en los cuales aún se continua investigando, ya que no se ha encontrado una solución definitiva. A la vez, hay muy poca información que involucre la combinación de los dos problemas [1], [2].

Debido a lo anterior, algunos trabajos se han enfocado en determinar cuál de estos dos efectos es el que más influye en el rendimiento de los sistemas de clasificación [3], [5], obteniendo en sus resultados que el traslape afecta más que el desbalance entre clases. Prati et al. en [5] emplearon conjuntos de datos generados artificialmente con diferentes relaciones de desbalances y traslapes para concluir que los problemas de rendimiento no son causados por el desbalance. Por su parte, Denil y Trappenberg en [3] demuestran que estos dos factores tienen efectos interdependientes, evaluando el comportamiento de máquinas de soporte vectorial (SVM) ante datos con diferentes relaciones tal como se planteó en trabajo de Prati. En [3] además se validó que estos sistemas de clasificación son insensibles al desbalance cuando

las clases son totalmente separables, pero cuando los dos problemas se unen las SVM no son tan efectivas a menos que se tenga una inmensa cantidad de datos de entrenamiento, lo cual aumenta la complejidad de los modelos. No obstante, cuando se elimina uno de los problemas o los dos, este mismo SVM no funciona correctamente, validando la hipótesis de que hay efectos que no son tenidos en cuenta cuando el desbalance y el traslape se trabajan separadamente.

En el trabajo realizado por Tang y Gao [6] también emplean SVM para su sistema de clasificación final, el cual mediante clasificación multi-modelos SVM combinados con algoritmos de vecinos cercanos kNN permite no solo clasificar si un dato corresponde a una clase o a otra, sino que también indica si ese dato se encuentra en una frontera donde existe traslape con lo que un observador puede tener una incertidumbre de la clase en la que se clasificó el dato. Los resultados finales de este clasificador fueron levemente superiores a los obtenidos por modelos SVM simples, a modelos basados en reglas kNN y el clasificador multi-modelo RIONA.

Otro trabajo que obtiene fronteras que delimitan una región de traslape es el realizado por Tang et al. en [2], en el que emplean un algoritmo de decisión suave que permite plantear varias posibilidades para datos dentro de una región con traslape, combinado con un algoritmo de detección de traslape entre clases. Los resultados de este trabajo fueron comparados con las técnicas de decisión denominadas “crisp decision”, que se basan en la regla de que el ganador se lleva todo, obteniendo resultados que la superan en un rango de 1% a 2% en incremento de acierto para el caso de poco traslape, pero para el caso de mayor cantidad de datos traslapados la mejora es en el rango de 0.01% a 0.03% o en algunas ocasiones los resultados son menores.

Por su parte, Batista y compañía en [7], [8] han abordado los problemas de desbalance y traslape desde el punto de vista de minimizar el primer problema a través de la implementación métodos para balancear el conjunto de datos de entrenamiento. Los métodos que analizan en sus trabajos aplican técnicas de sobremuestreo, para la clase con menor cantidad de datos, y de submuestreo para la clase con mayor información, obteniendo como resultados una mejora en la información de

las fronteras entre clases que a su vez incurre en la mejora del rendimiento de los modelos. Igualmente, comprobaron que los métodos de sobremuestreo superan a los resultados de los métodos de submuestreo.

García y compañía en [4] también emplean algoritmos basados en la regla kNN, y analizaron tres características de esta regla: la relación de desbalance total, el tamaño de la región de traslape y la relación local de desbalance en la región de traslape. Los resultados de su trabajo demostraron que el kNN depende más de los cambios en la relación de desbalance local en la región de traslape que de los cambios de tamaño de la región de traslape. A su vez encontraron que las clases más representadas en la región de traslape son mejor clasificadas en los métodos basados en aprendizaje global, mientras que las clases menos representadas en la región de traslape son mejor clasificadas por métodos locales.

Otros autores han trabajado en este problema con otra clase de clasificadores, como lo son del tipo de análisis discriminante. Han y compañía [9] trabajaron específicamente con el análisis discriminante escaso (SDA por su sigla del inglés Sparse Discriminant Analysis), extendiendo la teoría de los algoritmos de una sola tarea (Single-task StSDA) a la teoría multi-tarea (Multi-task SDA MtSDA) en la que se emplea una técnica de optimización cuadrática en la cual se usan valores de penalización, siendo en este caso matrices de penalización, para obtener resultados que faciliten la clasificación. La idea de emplear un algoritmo multi-tarea es porque esta clase de sistemas de aprendizaje usan información latente escondida entre clases, obteniendo buenos resultados en problemas de alta dimensionalidad y características de escasez de datos. En este trabajo se emplearon datos aleatorios de 11 bases de datos de Yahoo, las cuales se describen en [10], de la base de datos NUS-WIDE [11] y de la base de datos MSRA-MM2 [12]. Los resultados del sistema de clasificación MtSDA se compararon con un sistema de clasificación basado en SVM, otro que combina LDA con SVM, y un clasificador HSML, en los que en todos se obtuvo la mayor tasa de acierto con el clasificador propuesto, en las dos versiones planteadas, con una matriz identidad y una matriz de equicorrelación como matrices de castigo.

En la literatura, las máquinas de soporte vectorial y el análisis de discriminantes lineales Fisher son dos de las técnicas, supervisadas y no supervisadas, más usadas en reconocimiento de patrones y aprendizaje de máquinas aplicadas en clasificación. Y basados en los trabajos previos SVM y clasificadores basados en la regla kNN, son las técnicas más empleadas para abordar estos dos problemas.

Los discriminantes Fisher tienen gran aplicación debido a la simplicidad de cálculo y de ejecución, de manera que los algoritmos no requieren de gran costo computacional [13]. Las SVM tienen ligera mayor complejidad, pero a cambio se desempeñan generalmente mejor en condiciones de representación asimétrica de las clases en el conjunto de datos y permiten hacer ajustes finos mediante la modificación de parámetros de la SVM [14].

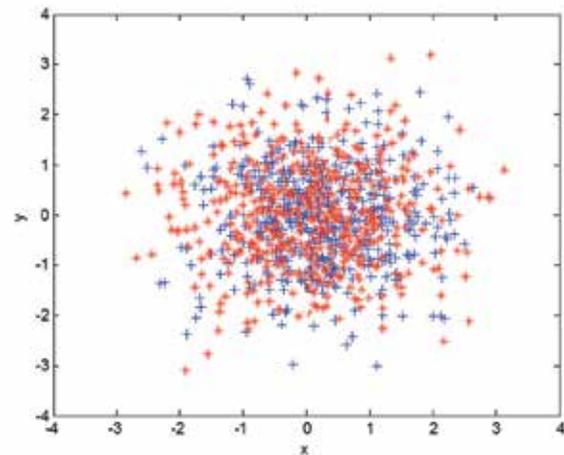
En la sección 2 se muestra el problema de clasificación de dos clases cuando hay traslape entre las clases. En la sección 3 se reseñan alternativas de solución a este tipo de problemas. En la sección 4 mostramos la aproximación a la solución de un problema de clasificación con variables débiles (weak features), haciendo énfasis en la capacidad de construcción de un sistema de clasificación fuerte a partir de parámetros débiles independientes o con cierto grado de independencia. En la sección 5 introducimos el algoritmo novedoso APCC (A Posteriori Confidence Classification) que permite asignar ponderación a cada parámetro débil de acuerdo con su respuesta y no únicamente de acuerdo con una ponderación previa. En la sección 6 mostramos los resultados de distintas técnicas usadas en este trabajo, así como un estudio de caso real con datos de clasificación de la expresión facial. Finalmente, en la sección 7 se elaboran las conclusiones de este trabajo.

2. EL PROBLEMA DE CLASIFICACIÓN CON TRASLAPE DE CLASES

Los sistemas de clasificación convencionales se basan en la presunción de que las clases son separables mediante un hiperplano de frontera. Esta separación se puede realizar en la misma dimensión de los datos originales o mediante la proyección supervisada de los datos a una dimen-

sión distinta en los cuales la distinción entre las clases es más sencilla. Sin embargo, esto es solo realizable cuando las clases están efectivamente bien representadas en el conjunto de muestra. En numerosos casos no es posible asegurar esta distinción, porque las clases representadas en el espacio n-dimensional presentan traslape, de manera que no existe un hiperplano de separación que permita garantizar adecuada tasa de clasificación acertada [15]. Un ejemplo de esto se plantea en la Fig. 1 en la que se presenta un problema en dos dimensiones de dos clases (clase 1 en color rojo y clase 2 en azul) las cuales se encuentran traslapadas y se hace difícil poder separar fácilmente las dos clases mediante una línea o un plano.

Fig. 1. EJEMPLO DE UN PROBLEMA DE CLASIFICACIÓN CON TRASLAPE



Fuente: autores.

Los clasificadores débiles tienen, sin embargo, una ventaja importante sobre los clasificadores convencionales. Un clasificador débil no tiene una gran correlación respecto de las clases, pero su eficiencia se basa en la combinación de un número de clasificadores tal que la fusión de un conjunto de clasificadores permite aumentar la tasa de acierto global. Adicionalmente, debido a la poca exigencia de cada clasificador, generalmente es suficiente usar un sistema de clasificación simple, reduciendo de esta forma la complejidad de la clasificación global y limitando posibles errores de *overfitting* y pobre generalización [16].

Un clasificador débil es aquel cuya capacidad de clasificación binaria es limitada y en muchos casos la tasa de clasificación exitosa es baja, apenas levemente superior a la clasificación aleato-

ria. Teniendo en cuenta que un clasificador binario cuya salida sea aleatoria independientemente de la entrada tiene en promedio una tasa de clasificación de 50%, un clasificador cuya tasa de acierto sea ligeramente superior es un clasificador no confiable en sí. Sin embargo, el uso de clasificadores débiles en un sistema de clasificación se basa en que la incorporación de un conjunto de clasificadores débiles permite construir un clasificador fuerte con tasas de acierto mucho mayores. La teoría de clasificación débil tiene la premisa de que los clasificadores individuales son independientes entre sí, lo cual no necesariamente es cierto en la realidad, pero incluso si los clasificadores individuales tienen cierta correlación entre sí, es posible garantizar la construcción de clasificadores más fuertes mediante fusión.

3. APROXIMACIONES A LA SOLUCIÓN DEL PROBLEMA DE CLASIFICACIÓN CON TRASLAPES DE CLASES

En una primera prueba, se usó un conjunto de dos clases con traslape entre clases relativamente bajo. La generación de este conjunto de datos fue basada en dispersión Gaussiana en dos dimensiones y el resultado de los conjuntos generados se muestra en la Fig. 2. Estos datos fueron modelados usando dispersión Gaussiana y media en los puntos (0,0) y (0,1) para las dos clases respectivamente. La clasificación se realizó usando FDA lineal (izquierda) y SVM-rbf (derecha).

Debido a la relativamente sencilla separación entre las clases, diversos métodos de clasificación binaria pueden ser usados, tales como métrica simple euclidiana, métrica Mahalanobis, FDA o SVM. El conjunto global fue dividido en 2 subconjuntos para entrenamiento y validación usando validación *10-folded*. En la sección 6 veremos los resultados.

En este caso la distancia entre clases y el poco traslape entre las muestras permite obtener tasas de clasificación adecuadas. La mayor parte de los errores son imposibles de evitar, por cuanto se trata de muestras situadas en la región correspondiente a la otra clase, de manera que en general los resultados son convenientes. En una segunda prueba se usó un conjunto de datos de naturaleza similar al anterior, pero el traslape entre clases fue aumentado al reducir la distancia entre las clases. Para ello se simuló dos clases con la misma dispersión gaussiana pero se redujo la distancia entre

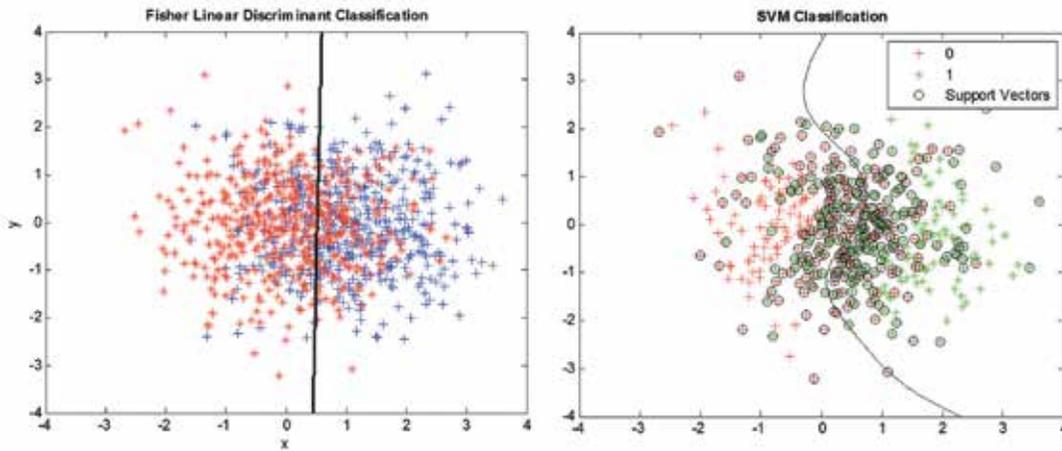
las clases al poner las medias en (0,0) y (0,0.5), respectivamente, y se hizo clasificación usando FDA lineal y SVM lineal, tal como se muestra en la Fig. 3.

Naturalmente, al reducir la distancia entre las clases se espera que la tasa de acierto de la clasificación se reduzca, de modo que el objetivo de la clasificación no es, ni podría obtener altas tasas de clasificación, sino al menos minimizar el error de clasificación.

Los anteriores experimentos se basan en la premisa de que las clases están al menos relativamente bien representadas en el conjunto de datos. Lamentablemente, en espacios *n*-dimensionales de alta dimensión es más difícil garantizar que las clases estén bien representadas, por cuanto el aumento de dimensión implica un aumento geométrico del número de puntos necesarios para representar adecuadamente una clase. Con el fin de ilustrar este inconveniente de manera visualizable, un nuevo conjunto de datos fue generado para simular sub representación de clases mediante la reducción del número de muestras por clase en un espacio 2-dimensional. Con el conjunto de datos obtenido se hizo entrenamiento y clasificación por distintos métodos. Los datos y las fronteras de clasificación son mostrados en la Fig. 4. En este caso una de las clases (en rojo) tiene considerablemente menor número de muestras, para simular la baja representación de una de las clases, y se hizo clasificación por FDA lineal (izquierda) y SVM-rbf (derecha).

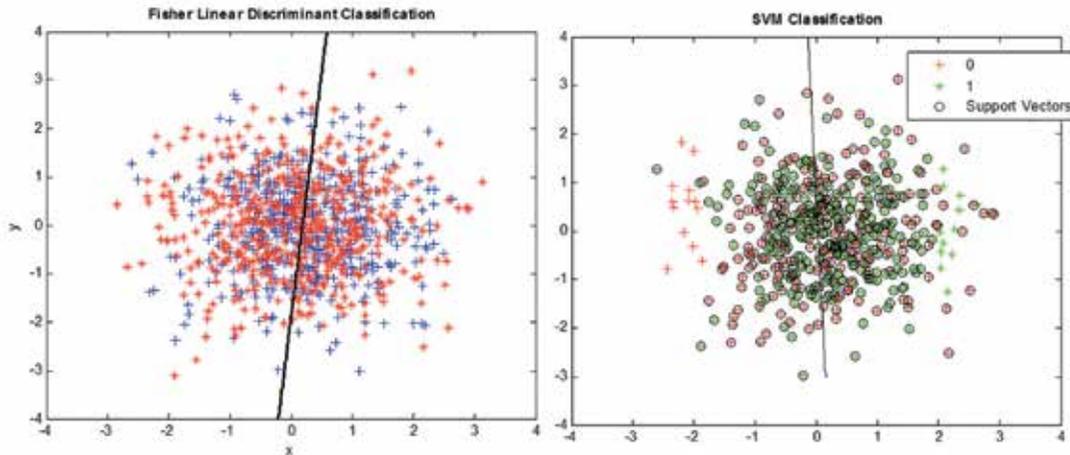
La pobre representación de las clases genera dificultades a los algoritmos de clasificación más sofisticados. Las máquinas de soporte vectorial con funciones de base radial (SVM-rbf) son particularmente propensas a error por pobre representación. Esto acontece porque las muestras seleccionadas como vectores de soporte no son necesariamente adecuadas para representar la generalidad de los datos de la clase y, de hecho, pueden incrementar el error por pobre generalización. Debido a ello las fronteras de clasificación usando SVM-rbf tienden a representar los datos de entrenamiento, pero la generalización con los datos de validación posiblemente es inadecuada. En oposición, las consecuencias de la pobre generalización con SVM-rbf no son tan notorias usando SVM lineales, de modo que en estos casos en particular la clasificación es más exitosa usando lineal SVM.

Fig. 2. EJEMPLOS DE CONJUNTOS DE DATOS CON DISPERSIÓN GAUSSIANA Y POCO TRASLAPE (FDA Y SVM-RBF)



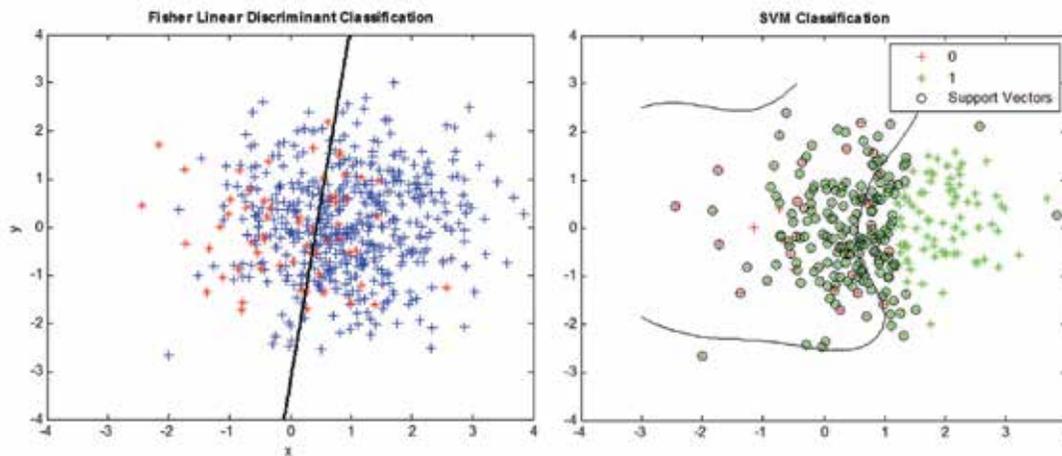
Fuente: autores.

Fig. 3. EJEMPLOS DE CONJUNTOS DE DATOS CON DISPERSIÓN GAUSSIANA Y MAYOR TRASLAPE (FDA Y SVM LINEAL)



Fuente: autores.

Fig. 4. EJEMPLOS DE CONJUNTOS DE DATOS DE DISPERSIÓN GAUSSIANA DESBALANCEADOS Y CON ALTO TRASLAPE (FDA Y SVM-RBF)



Fuente: autores.

La siguiente prueba fue realizada con datos con alto traslape y distinta dispersión. En este caso en específico, los datos de una clase están incrustados en un racimo dentro de la región general de los datos de la otra clase. En la Fig. 5 se visualizan los datos y las fronteras de clasificación usando diversas técnicas. Para ello los datos de una clase (rojo) son simulados con dispersión gaussiana con menor valor de σ , de manera que se encuentran dentro de la región general de la otra clase.

Las SVM-rbf tienen el mejor desempeño en este tipo de representación de datos, por cuanto la posibilidad de generar una frontera no lineal permite obtener una frontera de clasificación que se ajuste más adecuadamente a los datos reales. Los discriminantes Fisher lineales obtienen una frontera que naturalmente no permite discriminar los datos de manera tan precisa, pero al menos minimizan el error teniendo en cuenta las limitaciones de la frontera lineal. Las SVM lineales presentan un problema importante, que se reflejará más claramente en el uso de fusión de clasificadores débiles, por cuanto la frontera de clasificación si bien minimiza el error en la clasificación de una de las clases, el error de clasificación de la otra clase no es adecuado para la construcción de un clasificador fuerte. Este problema acontece porque las máquinas de soporte vectorial seleccionan un conjunto de vectores de soporte y obtienen un hiperplano de frontera que maximiza la distancia entre los vectores de

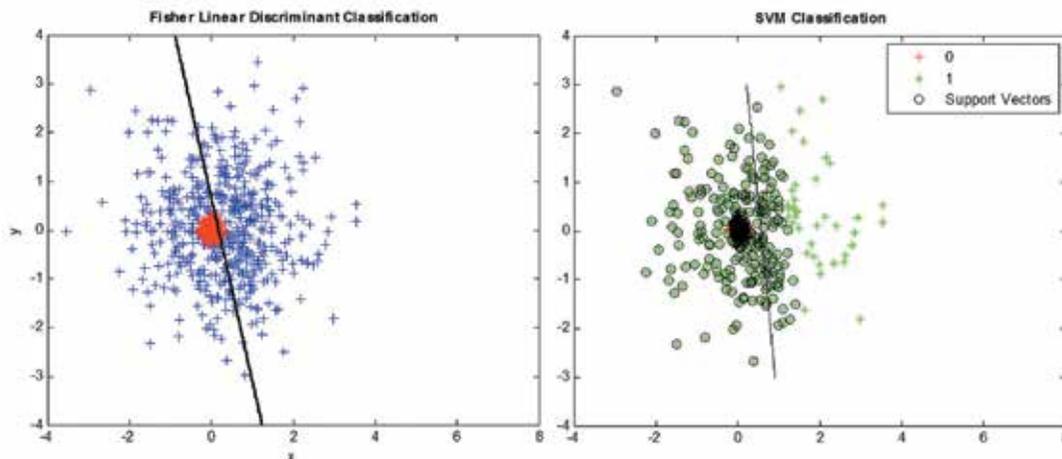
soporte y el hiperplano. Lamentablemente, en el caso de conjuntos fuertemente agrupados (clusters) embebidos en datos pertenecientes a la otra clase, la maximización de la distancia entre el hiperplano y los vectores de soporte hace que la frontera deba ser situada de modo que los datos de la clase II sean pobremente clasificados.

En la sección 6 mostraremos los resultados de clasificación por FDA, distancia euclidiana, SVM y SVM-rbf de los problemas reseñados previamente.

4. APROXIMACIÓN MEDIANTE CLASIFICACIÓN DÉBIL CON TRASLAPE DE CLASES

Algunos de los clasificadores mostrados en la sección anterior, particularmente cuando el traslape entre los datos es grande, son clasificadores débiles. Con el fin de ilustrar la fortaleza de los clasificadores débiles fusionados para obtener un clasificador fuerte, se usaron múltiples conjuntos débiles y mediante fusión simple se construyó un clasificador fuerte. Las primeras pruebas fueron realizadas con datos similares a los usados en la sección anterior, a saber: datos separados fácilmente diferenciables, datos pobremente diferenciables, datos sub-representados y datos embebidos en racimo en la región perteneciente a otra clase. La clasificación final fue obtenida por fusión simple mediante método *winner takes all*, según el siguiente pseudo algoritmo; Fig. 6.

Fig. 5 CONJUNTOS DE DATOS CON DIFERENTE DISPERSIÓN GAUSSIANA BALANCEADOS Y CON ALTO TRASLAPE (FDA Y SVM LINEAL)



Fuente: autores.

Fig. 6. PSEUDOALGORITMO OBTENIDO MEDIANTE WINNER TAKES ALL

```

Entrada
 $X_{k,c} = \{x_{1,k}, x_{2,k}, \dots, x_{N,k}\}$ ,  $C = \{C_1, \dots, C_N\}$ ,  $C_i \in \{1,2\}$ 
for i=1:N
  Obtener la salida de cada metaclassificador Class
   $Class_{k,c1,c2,t} \{x_{i,k}^{val}\} \rightarrow z_{i,k,c1,c2}$ 
  for k=1:K
    for cl=1:2
      for cl2=1:2
        if  $z_{i,k,c1,cl2} > 0$ 
           $Scl_{i,cl} = Scl_{i,cl} + 1$ 
        else
           $Scl_{i,cl2} = Scl_{i,cl2} + 1$ 
        end
      end
    end
  end
   $y_i = \operatorname{argmax}_c (Scl_{i,c})$ ,  $c = \{1,2\}$ 
end

```

Fuente: autores.

Donde $X_{k,c}$ es el conjunto completo de datos, $x_{i,k}^{val}$ es el conjunto de validación, $Scl_{k,c1,c2,t}$ son los metaclassificadores binarios entre la clase 1 y la clase 2, k es el índice correspondiente a cada clasificador débil (parámetro o *feature*), $Sclass$ es el vector de suma de la salida de los clasificadores binarios por parámetro y y_i es la clasificación final por *winner takes all*.

5. NIVEL DE CONFIANZA DE CLASIFICADORES DÉBILES INDIVIDUALES Y CLASIFICACIÓN POR CONFIANZA A POSTERIORI PONDERADA

Una alternativa para atenuar el efecto nocivo de la pobre generalización de los sistemas de clasificación sofisticados, cuando los datos de representación de cada clase no son ideales o el nivel de traslape entre clases es grande, es definir un nivel de confianza de clasificación individual. Con el fin de simular la presencia de diversos tipos de datos en clasificación débil, se generaron datos de dos clases de distinta naturaleza: con diversos grados de traslape, distancia entre clases, tipo de dispersión y agrupamiento de racimos. Para cada uno de los conjuntos generados se obtuvieron conjuntos de datos de entrenamiento, primera etapa de validación y segunda etapa de validación, sin repetir muestras entre los conjuntos. Los diversos datos fueron usados

para obtener clasificadores por métrica euclidiana, FDA, SVM lineal y SVM-rbf. Una vez obtenidos los clasificadores, se usó la primera etapa de validación para asignar un puntaje a cada clasificador de acuerdo con su capacidad de clasificación global. El clasificador elegido como metaclassificador débil fue obtenido mediante competencia simple todos contra todos, de manera que para cada conjunto de datos el clasificador seleccionado es aquel que permite mayor diferenciación entre los datos. Una vez hecho esto, en la etapa de validación final los clasificadores elegidos se probaron con los datos de la segunda etapa de validación.

La adición de un parámetro de confianza a cada clasificador permite que los clasificadores seleccionados como metaclassificadores débiles sean generalmente los idóneos para cada conjunto de datos, de manera que los resultados obtenidos muestran que la tasa de clasificación global es mucho más adecuada. Por otra parte, si bien la clasificación es exitosa en buena medida, la inclusión de un parámetro adicional mediante el uso de confianza a posteriori permite aumentar aún más la tasa de clasificación.

La inclusión de un parámetro de confianza a cada clasificador permite aumentar la tasa de clasificación. Sin embargo, es posible obtener una mejora notable en la clasificación si el parámetro de confianza no depende exclusivamente

de la capacidad individual de clasificación, sino además se tiene en cuenta el grado de confianza de clasificación de cada clasificador en función de su salida. Esto se muestra más adelante en la sección 6, tabla V, donde mostramos los resultados de clasificación de un estudio de caso

usando técnicas convencionales y nuestra técnica propuesta.

A este tipo de calificación de cada clasificador lo hemos denominado *A Posteriori Confidence Classification* (APCC). El pseudoalgoritmo de APCC es mostrado a continuación; Fig.7

Fig. 7. PSEUDOALGORITMO DE APCC

```

Entrada
 $X_k = \{x_{1,k}, x_{2,k}, \dots, x_{N,k}\}$ ,  $C = \{C_1, \dots, C_N\}$ ,  $C_i \in \{1,2\}$ 
Generación de datos para n-folded

for i=1:2
 $X_k$  es dividido aleatoriamente en 10 subconjuntos
end
Entrenamiento
for t=1:10
 $X_k^{tr}$ : subconjunto de entrenamiento
 $X_k^v$ : subconjunto de validación
for k=1:K
for cl=1:2
for cl2=1:2
if  $cl1 \neq cl2$ 
class 1= $X_{k,c1}^{tr}$ ,  $C_1 = cl$ 
class 2= $X_{k,c2}^{tr}$ ,  $C_2 = cl2$ 
Entrenar  $Class_{k,c1,c2,t}$ 
end
end
end
end
end

Input
 $X_{k,c} = \{x_{1,k}, x_{2,k}, \dots, x_{N,k}\}$ ,  $C = \{C_1, \dots, C_N\}$ ,  $C_i \in \{1,2\}$ 
Reliability per cell per output
for t=1:10
 $X_k^{tr2}$  subconjunto de entrenamiento 2
for i=1:N
for k=1:K
Obtener la salida de los metaclassificadores por conjunto
 $Class_{k,c1,c2,t}(x_{i,k}^{tr2}) \rightarrow z_{i,k,c1,c2}^{tr}$ 
end
end
 $P_{k,c1,c2,t} = P(x_{i,k}^{tr2} \in C_1 | z_{i,k,c1,c2}^{tr} > 0)$ 
 $P^j_{k,c1,c2,t} = P(x_{i,k}^{tr2} \notin C_1 | z_{i,k,c1,c2}^{tr} < 0)$ 
end
 $P_{k,c1,c2} = \frac{1}{10} \sum_t P_{k,c1,c2,t}$ 
 $P^j_{k,c1,c2} = \frac{1}{10} \sum_t P^j_{k,c1,c2,t}$ 

Validación
for t=1:10
 $X_k^v$  subconjunto de validación
for i=1:N
for k=1:K
Obtener la salida de los metaclassificadores por conjunto:
 $Class_{k,c1,c2,t}(x_{i,k}^v) \rightarrow z_{i,k,c1,c2}^v$ 
for cl=1:2
for cl2=1:2
if  $z_{i,k,c1,c2}^v > 0$ 
 $Sc_{i,k,c1} = c_{i,k,c1} + P_{k,c1,c2}$ 
if  $z_{i,k,c1,c2}^v < 0$ 
 $Sc_{i,k,c2} = c_{i,k,c2} + P^j_{k,c1,c2}$ 
end
end
end
end
The sample  $x_{i,k}^v$  is classified according to a winner takes all contest using the sum of the
outputs of the K features:
 $y_i = \operatorname{argmax}_c \sum_k Sc_{i,k,c}$ ,  $c \in \{1,2\}$ 

end
end
Construir la matriz de confusión

```

6. RESULTADOS Y ESTUDIO DE CASO

En la tabla I se muestran los resultados del problema de clasificación de dos clases, descrito en la figura 2, de separación binaria con bajo traslape. Los sistemas de clasificación usados son FDA, distancia euclidiana, SVM y SVM-rbf. Pese a la simplicidad de la clasificación por distancia euclidiana, los resultados de clasificación fueron mejores. La diferencia entre el promedio de clasificación con distancia euclidiana (0.8425) y con SVM (0.8255) no parece significativa. No obstante, dado el tamaño del conjunto de datos, esta diferencia es suficiente para mostrar una estadística diferente entre las dos pruebas usando t-test con 95% de intervalo de confianza.

TABLA I
SEPARACIÓN BINARIA ENTRE CLASES CON TRASLAPE BAJO

Clasificador	Clase A	Clase B
FDA	0.813	0.813
Euclidiana	0.838	0.847
SVM	0.827	0.824
SVM-RBF	0.829	0.820

El problema de separación binaria con alto traslape descrito en la Fig. 3 fue abordado usando las mismas técnicas que el problema anterior. Los resultados se muestran en la tabla II. Naturalmente, los resultados de clasificación son inferiores en conjunto, pero técnicas simples como distancia euclidiana permitieron obtener resultados mejores que con máquinas de soporte vectorial con discrepancia suficiente para mostrar diferencia estadística t-test.

TABLA II
SEPARACIÓN BINARIA ENTRE CLASES CON TRASLAPE ALTO

Clasificador	Clase A	Clase B
FDA	0.782	0.776
Euclidiana	0.793	0.795
SVM	0.781	0.785
SVM-RBF	0.779	0.778

Al desbalancear los conjuntos de datos, tal como se muestra en la Fig. 4, los algoritmos de clasificación tienen una respuesta distinta, tal como se muestra en la tabla III. Debido a las funciones de castigo usadas, orientadas a mejorar la tasa global de clasificación, la clase B es clasificada correctamente con más frecuencia, debido al mayor número de muestras. En este caso no hay diferencia

estadística significativa entre clasificación por FDA, distancia euclidiana y SVM, pero el uso de SVM-rbf tiende a sobreajustar los datos salvo modificaciones finas a los parámetros de cada SVM, de manera que mejora la clasificación de la clase B, pero empeora notablemente la clasificación global.

TABLA III
SEPARACIÓN BINARIA ENTRE CLASES CON DESBALANCE Y ALTO TRASLAPE

Clasificador	Clase A	Clase B
FDA	0.645	0.876
Euclidiana	0.613	0.892
SVM	0.613	0.884
SVM-RBF	0.463	0.908

En la tabla IV mostramos los resultados de clasificación con un conjunto de datos parcialmente embebido dentro de otro conjunto, como en la Fig. 5. Este problema ilustra con mayor claridad las desventajas de sistemas sofisticados como SVM en un problema de parámetros débiles. La fusión de parámetros débiles con clasificadores simples FDA y euclidianos permitió codificación perfecta de la clase A y alrededor de 82% de la clase B. En cambio, SVM y SVM-rbf tuvo clasificación nula de la clase B. Esto ocurre debido a los hiperplanos de separación creados, similares a los mostrados en la Fig. 5. Si bien los hiperplanos de separación por SVM o SVM-rbf fueron adecuados para discriminar la clase A, no aportan información suficiente para separar la clase B, de manera que la fusión, por *winner takes all*, tiene tasa nula de acierto para la clase B.

TABLA IV
SEPARACIÓN BINARIA ENTRE CLASES BALANCEADAS PERO CON DIFERENTE DISPERSIÓN Y ALTO TRASLAPE

Clasificador	Clase A	Clase B
FDA	1.000	0.812
Euclidiana	1.000	0.824
SVM	1.000	0.000
SVM-RBF	1.000	0.000

Para validar y simular la metodología de clasificación desarrollada en este trabajo se creó un conjunto de datos de 101 parámetros débiles de 2 variables. Las variables son de naturaleza gaussiana de alta dimensión (dimensión 30) con niveles aleatorios de dispersión, traslape y representación de cada clase binaria. De esta forma se garantiza que diversos tipos de representación de datos son simulados. La validación fue realizada usando una

extensión de *10-folded cross validation*. El conjunto total de datos fue dividido en 3 grupos: un grupo de entrenamiento de los clasificadores débiles euclidianos, FDA y SVM, otro grupo de entrenamiento de los metaclasificadores débiles por APCC y un grupo de validación. Esta metodología permite separar el entrenamiento de los clasificadores débiles euclidianos, FDA y SVM de los metaclasificadores APCC, de manera que se reduce la posibilidad de pobre generalización en el entrenamiento. La matriz de confusión obtenida se presenta en la tabla V.

Los resultados confirman que variables con fuerte traslape y pobre correlación con las clases ofrecen resultados inferiores de fusión de clasificación cuando se usan métodos sofisticados de clasificación, tales como SVM. En contraste, metodologías simples de clasificación como métricas euclidianas y clasificadores FDA ofrecen mayor generalización de los datos. Existen alternativas para mejorar el desempeño de los clasificadores basados en SVM, tales como la manipulación individual de los parámetros de las SVM, generación de funciones de castigo y empleo o no de kernels en función de los datos por representar. No obstante, estas modificaciones individuales incrementan costos de entrenamiento y validación y se aumenta el riesgo de introducir overfitting indeseado, por cuantos se podría obtener clasificadores que aprendan las clases en función de los datos, con baja generalización.

Por último, se diseñó una función de clasificación como fusión de los clasificadores obtenidos en las pruebas anteriores. Debido a que cada metaclasificador tiene un puntaje de confianza dado por el algoritmo APCC, los datos de los 4 metacla-

sificadores por parámetro débil fueron fusionados. La fusión es determinada en función del parámetro de confianza de cada metaclasificador. De esta forma se pretende que cada parámetro débil sea evaluado con mayor puntuación por los metaclasificadores con mayores parámetros de confianza y así mejorar la clasificación global. Los resultados obtenidos se muestran en la tabla VI.

Estos resultados muestran que el desempeño mejora notablemente al usar la fusión de los metaclasificadores individuales. Adicionalmente, es de reseñar que un incremento de aproximadamente 1,5% en la clasificación global puede parecer pequeño, pero cuando la tasa de clasificación acertada es cercana a 100%, es más difícil mejorar el desempeño total, de manera que incluso valores como 1,5% son una muestra determinante de la mejoría de la clasificación.

Estas técnicas también fueron usadas en el diseño de clasificación multiclase en un problema de 7 clases, para reconocimiento de la expresión facial usando parámetros TPOEM (*Temporal Pattern of Oriented Edge Magnitudes*) [17] y VPOEM (*Volumetric Patterns of Oriented Edge Magnitudes*) [18] con clasificadores ponderados a partir de parámetros débiles. Los resultados obtenidos usando clasificación Mahalanobis, FDA convencional y fusión de clasificadores siguiendo el protocolo mostrado en este artículo. La validación fue realizada usando conjuntamente *10-folded cross validation* y *bootstrapping* y los resultados se muestran en la tabla VII.

Si bien una mejora de 85,3% con clasificación convencional Mahalanobis a 89,67% con fusión de clasificadores y APCC no parece ser significa-

TABLA V
MATRIZ DE CONFUSIÓN ESTUDIO DE CASO

Clasificadores a	Euclidianos		FDA		SVM		SVM-RBF	
	Clase 1	Clase 2	Clase 1	Clase 2	Clase 1	Clase 2	Clase 1	Clase 2
Clase 1	96.23%	3.77%	96.81%	3.19%	88.73%	11.27%	81.77%	18.23%
Clase 2	3.48%	96.52%	3.26%	96.74%	10.94%	90.06%	14.49%	82.51%

TABLA VI
MATRIZ DE CONFUSIÓN EMPLEANDO APCC

Clases	Clase 1	Clase 2
Clase 1	98,16%	1,84%
Clase 2	1,75%	98,25%

TABLA VII
RESULTADOS EN UN PROBLEMA DE 7 CLASES USANDO PARÁMETROS TPOEM

	Mahalanobis	FDA	SVM	Fusión APCC
Clasificación	85,3%	85,34%	83,63%	89,67%

tiva, mejoras pequeñas en tasas de clasificación cuando la clasificación se acerca a 100% representan mayor dificultad. En este caso, la tasa de error bajó de 14,7% a 10,33%, lo que corresponde a una disminución significativa del error.

7. CONCLUSIONES

En este trabajo se pudo verificar que metodologías sofisticadas de clasificación como SVM y SVM con funciones de base radial tienen desempeño deficiente cuando las clases binarias tienen representación inadecuada, específicamente por fuerte traslape entre las clases y número reducido de muestras por clase. En estas condiciones, clasificadores más simples pueden tener resultados mejores si la clasificación global se realiza como fusión de clasificadores débiles. Por otra parte, al introducir parámetros de confianza a posteriori APCC, se obtiene un criterio más fuerte de credibilidad de cada clasificador débil a partir de la evaluación de la salida. De esta forma, la introducción del algoritmo APCC permitió mejoras notables en la clasificación, incluso en sistemas representados por un número relativamente reducido de metaclasificadores débiles (101, en oposición a miles o decenas de miles de clasificadores débiles usados en sistemas, tales como detección de rostro por AdaBoost).

Si bien la clasificación simple por métrica euclidiana o por FDA tiene resultados individuales limitados, la fusión de clasificadores permite obtener resultados globales mucho mejores. En cambio, métodos de clasificación más sofisticados como SVM lineales o SVM-rbf tienen problemas de clasificación al fusionar los datos. Estos inconvenientes si bien parecen extraños, son fácilmente explicables teniendo en cuenta las limitaciones de las máquinas de soporte. Las restricciones de distancia a los vectores de soporte impuestas en la obtención del hiperplano de separación en un SVM hacen que en algunos de los casos mostrados la tasa de acierto para la clase II pueda ser menor que el 50%. Como consecuencia, el resultado de la fusión de clasificadores ocasiona que en estos

casos las muestras sean clasificadas como pertenecientes a la clase I, de modo que el error de clasificación de la clase I es 0%, pero el error de clasificación de la clase II es 100%, lo que constituye un clasificador inútil en términos prácticos.

En la metodología de este trabajo los parámetros débiles tienen una tasa de clasificación de entre 52 y 56% individualmente. Es decir, cada clasificador binario es sin duda un clasificador débil, por cuanto la tasa de clasificación acertada es apenas ligeramente superior que la de clasificación aleatoria. Naturalmente, si en el conjunto de datos hubiese parámetros cuya clasificación individual fuese superior, la tasa de acierto mejoraría notablemente, pero habría sesgo en la metodología. Debido a ello se garantizó que ningún metaclasificador tuviese tasa de acierto superior a 56% en ninguna de las pruebas, y con ello se pudo verificar la validez del algoritmo propuesto.

Si bien este trabajo alcanza los objetivos planteados de desarrollo de sistemas de metaclasificación que permitan el tratamiento de parámetros débiles con fuerte traslape y baja representación, existen implementaciones que permiten mejorar aún más los resultados obtenidos. Las técnicas de fusión implementadas pueden ser mejoradas mediante el uso de arquitecturas de árboles bayesianos en vez de fusión simple, de manera que en función de la salida algunos parámetros sean descartados en el sistema de clasificación. Adicionalmente, todo el trabajo fue desarrollado con el uso de variables gaussianas, que si bien representan buena parte de los datos obtenidos en un sistema real, en diversas ocasiones la dispersión de las variables medidas no es necesariamente gaussiana y en estos casos es posible que el uso de kernels de transformación previos a la etapa de entrenamiento arroje resultados mejores.

8. REFERENCIAS

- [1] H. Xiong, J. Wu, and L. Liu, "Classification with Class Overlapping: A Systematic Study," in *The 2010 International Conference on E-Business Intelligence*, 2010, pp. 491-497.

- [2] W. Tang, K. Z. Mao, L. O. Mak, and H. W. Ng, "Classification for overlapping classes using optimized overlapping region detection and soft decision," in *Information Fusion (FUSION), 2010 13th Conference on*, 2010, pp. 1-8.
- [3] M. Denil and T. Trappenberg, "Overlap versus Imbalance," in *Advances in Artificial Intelligence*, 2010, pp. 220-231.
- [4] V. García, R. A. Mollineda, and J. S. Sánchez, "On the k-NN performance in a challenging scenario of imbalance and overlapping," *Pattern Anal. Appl.*, vol. 11, no. 3-4, pp. 269-280, Sep. 2007.
- [5] R. C. Prati, G. E. A. P. A. Batista, and M. C. Monard, "Class imbalances versus class overlapping: an analysis of a learning system behavior," in *MICAI 2004: Advances in Artificial Intelligence*, 2004, pp. 312-321.
- [6] Y. Tang and J. Gao, "Improved classification for problem involving overlapping patterns," *IEICE Trans. Inf. Syst.*, vol. 90, no. 11, pp. 1787-1795, 2007.
- [7] G. E. Batista, R. C. Prati, and M. C. Monard, "Balancing strategies and class overlapping," in *Advances in Intelligent Data Analysis VI*, 2005, pp. 24-35.
- [8] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explor. Newsl.*, vol. 6, no. 1, p. 20, Jun. 2004.
- [9] Y. Han, F. Wu, J. Jia, Y. Zhuang, and B. Yu, "Multi-Task Sparse Discriminant Analysis (MtSDA) with Overlapping Categories," in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*, 2010, pp. 469-474.
- [10] S. Ji, L. Tang, S. Yu, and J. Ye, "Extracting shared subspace for multi-label classification," in *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*, 2008, pp. 381-389.
- [11] T. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, "NUS-WIDE: A Real-World Web Image Database from National University of Singapore," in *Proceedings of the ACM International Conference on Image and Video Retrieval*, 2009, p. 48.
- [12] M. Wang, L. Yang, and X. Hua, "MSRA-MM: Bridging Research and Industrial Societies for Multimedia Information Retrieval," *Microsoft Res. Asia, Tech. Rep.*, pp. 1-14, 2009.
- [13] S.-J. Kim, A. Magnani, and S. P. Boyd, "Robust fisher discriminant analysis," in *Advances in Neural Information Processing System*, 2005, vol. 1, pp. 659-666.
- [14] S. Abe, *Support vector machines for pattern classification*. Springer, 2010.
- [15] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121-167, 1998.
- [16] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *J. Mach. Learn. Res.*, vol. 11, no. 1, pp. 2079-2107, 2010.
- [17] N.-S. Vu and A. Caplier, "Enhanced patterns of oriented edge magnitudes for face recognition and image matching," *Image Process. IEEE Trans.*, vol. 21, no. 3, pp. 1352-1365, 2012.
- [18] E. Silva, C. Esparza, and Y. Mejía, "POEM-based Facial Expression Recognition, a New Approach," in *Image, Signal Processing, and Artificial Vision (STSIVA), 2012 XVII Symposium of*, 2012, pp. 162-167.