

Teoría para Reconocimiento de Patrones Colaborativo

Dr. Argelio Victor de la Cruz Rivera¹, M. en C. María de los Ángeles Alonso Lavernia¹

¹ Centro de Investigación en Tecnologías de Información y Sistemas
Universidad Autónoma del Estado de Hidalgo (CITIS-UAEH)
argelioc@uaeh.reduaeh.mx, marial@uaeh.reduaeh.mx

Resumen

Se presentan los resultados preliminares obtenidos en el desarrollo de una teoría que permite la integración de varios métodos de Reconocimiento de Patrones Supervisado.

La solución propuesta consiste, en construir una Base de Conocimientos a partir de los datos contenidos en la Información de Aprendizaje, la cual describe las características de las clases existentes. Dado que esta transformación se ejecuta utilizando una estructura semejante a la de un sistema basado en conocimientos, este resultado permite construir sistemas que se nutran con experiencias humanas y datos provenientes del problema, además de posibilitar el uso colaborativo de diversos clasificadores.

Para realizar lo anterior, se desarrolla un procedimiento de descomposición de la información además de adecuaciones a varios métodos, para permitir su utilización bajo los mismos criterios. Algunos de estos métodos son: la regla de los k vecinos más cercanos, la regla del análisis global por grupo, la regla del análisis distinguido, árboles de decisión, etc.

Palabras Claves: Reconocimiento Supervisado, Base de Conocimientos, Clasificación Colaborativa.

1. Introducción

Como es conocido, los métodos de Reconocimiento de Patrones pueden ser divididos en dos grupos fundamentales: Reconocimiento Supervisado y Reconocimiento No Supervisado, aunque algunos autores también incluyen problemas de selección de variables y variantes de los anteriores [17], [19].

En ambos casos, se parte de suponer que se tiene una matriz de datos, que denominaremos M , la cual está formada por " n " filas y " m " columnas, donde las filas representan los individuos bajo estudio (también llamados objetos o entes, que denotaremos como I_i) y las columnas los parámetros o variables que han sido medidos a esos individuos (que se denotarán como V_j).

Los métodos de Reconocimiento No Supervisado tratan de resolver el problema de encontrar una clasificación adecuada para los individuos en N_c clases o grupos, mientras que los métodos de Reconocimiento Supervisado, parten de suponer que los individuos de la matriz M , ya se encuentran clasificados en N_c clases y el problema a resolver consiste en: clasificar otro conjunto de individuos, cuya pertenencia a las clases de M es desconocida. Es común, denominar como Matriz de Aprendizaje o Información de Aprendizaje a la matriz M y denominar como patrones a los individuos [6], [7], [8], [17].

Son diversos los trabajos que se han realizado en esta disciplina, sin embargo, continúan presente diversos problemas a la hora de su utilización práctica:

1. La aplicación de varios métodos para un mismo problema, puede provocar aparición de varias soluciones y es muy difícil encontrar una vía formalizada para tomar una decisión con todos esos resultados.
2. Ante un problema real no es posible determinar, a priori, cuál o cuáles métodos utilizar para garantizar que los resultados sean correctos.

En Reconocimiento de Patrones, la idea de combinar la decisión de varios clasificadores ha sido explorada con vistas a resolver estos problemas. Diversos métodos han sido usados para combinar los resultados de distintas redes neuronales [5], [16], [19], [21], utilizan árboles de decisión [2], [3] u otros modelos [1], [4], [12], [14].

En la mayoría de los casos se trata la combinación de modelos derivados de múltiples versiones del mismo método de aprendizaje o reconocimiento y por lo general se parte los resultados brindados por cada uno de ellos de forma independiente, para buscar una manera de combinarlos y obtener un único resultado [15], [18], [22].

Las razones para intentar combinar los resultados de múltiples clasificadores son fuertes, puesto que cada método puede representar aspectos diferentes de un problema de los datos de entrada como tal.

En el presente trabajo, se proponen soluciones a la utilización combinada clasificadores supervisados, pero mediante la integración de los métodos a arquitectura única, en lugar de la integración de sus resultados como tal.

Sin embargo, los resultados obtenidos no se centran únicamente en esta dirección, que la aproximación que se propone, permite el diseño de sistemas capaces de abordar la solución de problemas complejos de diagnóstico, pronóstico o toma de decisiones general, mediante la combinación de datos provenientes de una Información Aprendizaje y conocimientos provenientes de expertos humanos en el tema de forma colaborada.

2. Uso Combinado de Datos y Conocimientos

En la mayoría de los problemas reales de clasificación, diagnóstico, pronóstico o toma de decisiones en general, se pueden identificar dos fuentes de información fundamentales, que pueden ser usadas para buscar las soluciones:

- Casos conocidos representados a través de datos
- Experiencias Humanas

No se puede asegurar que una u otra información sea más importante, puesto que depende de cada caso en particular, y ambas pueden ser importantes fundamentales.

Para poder construir un sistema que sea capaz de manejar ambas fuentes de información es necesario encontrar formas de unificación de sus estructuras de representación.

La solución propuesta en este trabajo, consiste en la creación de nuevas estructuras migración de los datos hacia una estructura de Base de Conocimientos. Con esto se logra permitir la combinación de clasificadores, y se brinda una herramienta que permita el combinado de datos y conocimientos mediante estructuras híbridas [20].

En resumen, se brindan soluciones generales a problemas como los siguientes:

- Se posibilita el uso de diferentes criterios para Reconocimiento de Patrones de forma tal, que los resultados que se brinden sean obtenidos producto de la combinación los diversos criterios empleados. Cada método puede tener un aporte diferente resultado final que se brinda.

- Se permite el uso combinado de experiencias (conocimientos humanos) y casos conocidos (datos) para la solución de un mismo problema.
- Se brinda una herramienta computacional que permite el diseño de sistemas interactivos para Reconocimiento de Patrones, con posibilidades para justificar su comportamiento y los resultados que brinde [9], [10], [11].

Base de Conocimientos de la Información de Aprendizaje

En la teoría que será tomada como base [9], [10], [11], el término Base de Conocimientos (BC) se interpreta, hablando de forma global (no se pretende construir una definición formal) como la unión de dos componentes fundamentales:

- 1) Un conjunto formado por *hechos o conceptos* provenientes de un dominio del conocimiento específico.
- 2) Un conjunto de *relaciones* entre los elementos del conjunto anterior.

Una Forma de Representación de Conocimientos será considerada como toda aquella estructura teórica que permita expresar de alguna forma, los elementos de alguno de los conjuntos anteriores y a su vez posibilite procesar la información que ellos representada.

El trabajo que se presenta se enfoca hacia los problemas de Reconocimiento Supervisado, por tanto, la información de partida será la Matriz de Aprendizaje de algún problema particular y el objetivo es construir una BC, que describa las características de cada clase de esta información por separado.

Construcción de Conceptos

La vía más simple que utilizaremos para representar los elementos del primer conjunto una BC, es a través de las llamadas **Proposiciones**.

Una Proposición constituye una estructura con diferentes componentes, que permiten representar conceptos de diversa forma [9], [10], [11] pero sólo analizaremos los aspectos generales relacionados con la utilización que se le dará en el presente trabajo.

Una situación del mundo real, se puede representar en la computadora mediante una proposición, si a ésta se le puede asociar un valor de seguridad o certidumbre.

Las componentes básicas que utilizaremos para representar una proposición serán: textos identificadores, para expresar en lenguaje natural la situación que se desea representar, valor de certidumbre asociado y atributos naturales.

El valor de certidumbre $C(P_j)$ asociado a una proposición denotada como P_j , es un número del intervalo $[-1,1]$, donde 1 indica *verdad absoluta*, -1 falsedad absoluta, (lo contrario de lo anterior), y 0 *desconocimiento total*, es decir, que no se sabe. El resto de los valores se encuentran asociados a gradaciones de la veracidad.

Los atributos naturales pueden ser: pregunta, objetivo o intermedio, pero estos quedan definidos una vez que se introduce el segundo conjunto. Las proposiciones pregunta representan información conocida sobre el problema que debe ser suministrada como dato entrada. Los objetivos constituyen los resultados que se brindarán, y los intermedios representan subobjetivos dentro del problema que también deberán ser obtenidos.

4.1. Construcción de Proposiciones Objetivo

Como los problemas de Reconocimiento Supervisado se enfocan a clasificar un objeto una o varias clases entre las existentes en la Matriz de Aprendizaje, entonces se pueden definir las proposiciones objetivo como se expresa a continuación:

Si suponemos que existen N_c clases reconocidas en la matriz M , entonces se puede definir el texto general de las proposiciones objetivo como sigue:

“El individuo o patrón a reconocer, denotado como I_i , se clasifica en la clase l perteneciente a la Matriz de Aprendizaje”, donde I_i representa un individuo caracterizado por el mismo conjunto de variables que los miembros de la Matriz de Aprendizaje, pero que no se encuentra en ella (se denotará $I_i \notin M$) y $l \in \{1, 2, \dots, N_c\}$.

Si denotamos además, como O_l a la proposición objetivo l , entonces los valores de certidumbre asociados a O_l se interpretarían como sigue: si $C(O_l) > 0$ ello indicará que existen evidencias a favor de la clasificación o pertenencia del individuo I_i a la clase l claro está que mientras mayor sea este valor mayor seguridad habrá al respecto, pero por el contrario, $C(O_l) < 0$ entonces ello indica que existen razones para considerar que no pertenece a la clase l y $C(O_l) = 0$ indica que no se sabe nada acerca de la pertenencia del individuo a la clase.

Se puede entonces plantear que la primera tarea de un sistema utilizando la teoría propuesta, consistirá en lo siguiente:

“Dado una Matriz de Aprendizaje M y un individuo $I_i \notin M$, obtener los grados de certidumbre asociados a las N_c proposiciones objetivo derivadas de M , que expresan seguridad que se tiene sobre la pertenencia de I_i a cada clase”.

De lo anterior se evidencia la necesidad de definir como ejecutar este proceso y para ello establecer como relacionar la información de partida con los resultados expresados.

Por otra parte, no es evidente cuáles proposiciones preguntas o intermedios se generan, puesto que ello va a depender de las relaciones (elementos del segundo conjunto en BC) que se establezcan, de los métodos utilizados y de la naturaleza de los datos.

5. Construcción de Relaciones

Las relaciones entre los conceptos de una BC constituyen la vía utilizada para ejecutar los procesos de inferencia que permiten dar solución al problema planteado.

En nuestro caso, estas relaciones se conciben como los diferentes métodos de reconocimiento a ser aplicados, pero dado los objetivos y extensión del presente trabajo nos concentraremos en estudiar sólo algunos casos generales, que permiten construir una arquitectura sobre la que se pueden incorporar otros métodos o variantes.

5.1. Método de Descomposición de la Información

Como su nombre lo indica este método basa su funcionamiento en buscar la descomposición de la información para que el proceso de clasificación se ejecute en pasos. Comenzaremos el análisis planteando algunas definiciones para las variables numéricas.

5.1.1. Proposiciones de Tipo 1

Definición 1.- Una variable numérica j para el individuo I_i con valor V_{ij} es semejante a la variable j de la clase l si cumple que se encuentra dentro del intervalo $[\text{Min}j_l, \text{Max}j_l]$ donde $\text{Min}j_l$ denota al mínimo de la variable j para la clase l y $\text{Max}j_l$ el máximo de la variable j para la clase l .

Si denotamos como m_n a la cantidad de variables numéricas de la matriz, entonces utilizando esta definición se puede construir un conjunto de proposiciones preguntas precisamente con m_n elementos de la siguiente forma:

Proposición general Tipo1: La variable numérica j para el individuo I_i es semejante a variable j de la clase l . Se denota $PTipo1'_{ji}$.

Como se puede observar, se obtienen m_n proposiciones con atributo pregunta, de este planteamiento general.

1.1.1. Cálculo de la certidumbre asociada a las proposiciones Tipo 1

Denotaremos como $C(PTipo1'_{ji})$ la certidumbre asociada a las proposiciones Tipo1 y consideraremos que la zona alrededor de los valores máximo y mínimo

la variable pueden ser dudosos, para lo cual se introduce un Factor $F \in [0,1]$ que permitirá definir una función, para cada caso particular, en la que se represente la parte típica del intervalo, de la siguiente forma:

Si $F = 0$ ello indica que $C(PTipo1'_{ji}) = 1$ sólo para $V_{ij} = (Minj_l + Maxj_l) / 2$, es decir, cuando el valor de la variable j para el individuo I_i (V_{ij}) este exactamente en el centro del intervalo.

Para $F = 1$ se indica que si $V_{ij} \in [Minj_l, Maxj_l]$ entonces $C(PTipo1'_{ji}) = 1$ y si $V_{ij} \notin [Minj_l, Maxj_l]$ entonces $C(PTipo1'_{ji}) = -1$.

Los valores entre 0 y 1 son situaciones intermedias que permiten ampliar o disminuir la zona de dudas alrededor del mínimo y máximo. En la figura 1 se muestra un diagrama de función para un caso donde $F = 1/2$.

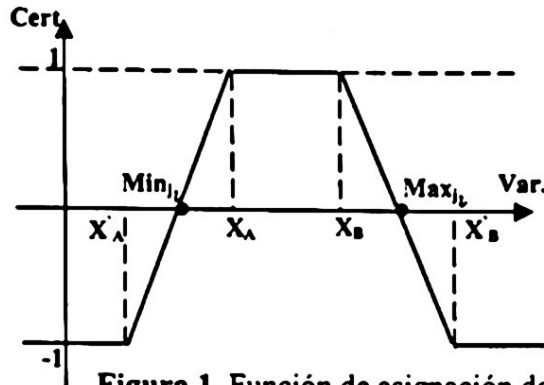


Figura 1. Función de asignación de certidumbre para $F=1/2$.

5.1.2. Reglas de Tipo 1

A partir de las proposiciones Tipo1 se obtiene información sobre la pertenencia del individuo a cada clase para una variable en particular como sigue:

Definición 2.- Si la variable numérica j para el individuo I_i , es semejante a la variable j de la clase l con certidumbre 1 ($C(PTipo1'_{ji}) = 1$) entonces el individuo I_i pertenece a la

clase l para la variable j con valor de certidumbre: $W_l = \frac{Nc - 1 - n_{jk}}{Nc - 1 + n_{jk}}$ donde Nc denota

la cantidad de clases y n_{jk} la cantidad de clases para las cuales $V_{ij} \in [Minj_k, Maxj_k]$, $k \in \{1, 2, \dots, Nc\} \setminus \{l\}$.

También se puede considerar que si la variable numérica j para el individuo I_i , no es semejante a la variable j de la clase l con certidumbre 1 ($C(PTipo1'_{ji}) = -1$) entonces constituye una proposición verdadera que el individuo I_i no pertenece a la clase l para j .

Se puede construir entonces, un conjunto de proposiciones de Tipo2 como sigue:

Proposición general Tipo2: El individuo I_i pertenece a la clase l para la variable numérica j . Se denota $PTipo2'_{ij}$.

Y como se observa en la definición 2 existe una relación entre las proposiciones Tipo1 las Tipo2, que puede ser expresada mediante una Regla de Producción como sigue:

Regla Tipo 1 (RT1): $PTipo1'_{ji} \Rightarrow PTipo2'_{ij} (W_l, -1)$;

Esta regla significa lo siguiente: Si se obtiene con completa seguridad que la variable numérica j para el individuo I_i es semejante a la variable j de la clase l ($C(PTipo1'_{ji}) = 1$)

entonces esto contribuye con W_i a la seguridad que se tiene sobre la pertenencia del individuo I_i a la clase l para la variable numérica j , pero si esta premisa no se satisface con completa seguridad ($C(PTipo1'_{ji}) = -1$) entonces contribuye con -1 .

Claro está, que si la premisa toma algún valor no extremal entonces la contribución debe ser re-calculada. Esto puede ser obtenido de diversas formas, pero una vía inmediata que garantiza obtener valores proporcionales al incremento o decremento de los valores asociados a la premisa es la siguiente: si $C(PTipo1'_{ji})$ denota la certidumbre asociada a $PTipo1'_{ji}$, y ésta es positiva, entonces la contribución será $C(PTipo1'_{ji}) * W_i$ y si es negativa entonces contribuirá con $-C(PTipo1'_{ji})$.

Como se puede observar, las proposiciones de Tipo2 son intermedios ya que no representan datos de entrada ni los objetivos definidos en el epigrafe 4.1. Veamos entonces como construir las reglas que permiten obtener los objetivos.

5.1.3. Reglas de Tipo 2

Definición 3.- Si el individuo I_i pertenece a la clase l para la variable numérica j con completa seguridad ($C(PTipo2'_{ij}) = 1$) entonces esto contribuye en $C(PTipo2'_{ij})/m$ (m denota la cantidad de variables) a la certidumbre global de la pertenencia de este individuo I_i a la clase l . $C(PTipo2'_{ij})$ denota la certidumbre de las proposiciones Tipo2.

Como se puede observar aquí se plantea una relación entre las proposiciones de Tipo2 y los objetivos, que se denominarán de Tipo3 (se denotarán $PTipo3'_i$), lo cual puede ser expresado a través de la siguiente regla general:

Regla Tipo 2 (RT2): $PTipo2'_{ij} \Rightarrow PTipo3'_i$ ($1/m, -1/m$); lo cual significa lo siguiente:

Si se obtiene con completa seguridad que el individuo I_i pertenece a la clase l para la variable numérica j ($C(PTipo2'_{ij}) = 1$) entonces esto contribuye en $1/m$ a la seguridad que se tiene sobre la pertenencia de I_i a la clase l , y si por el contrario se obtiene que $C(PTipo2'_{ij}) = -1$, entonces esto contribuye con $-1/m$.

Se puede observar que este criterio define aportes semejantes para todas las variables a la clasificación, sin embargo, es común en los problemas reales que unas variables sean más importantes que otras, de aquí que de ser conocida la importancia o peso de las variables, se puede utilizar esta información para ponderar las contribuciones.

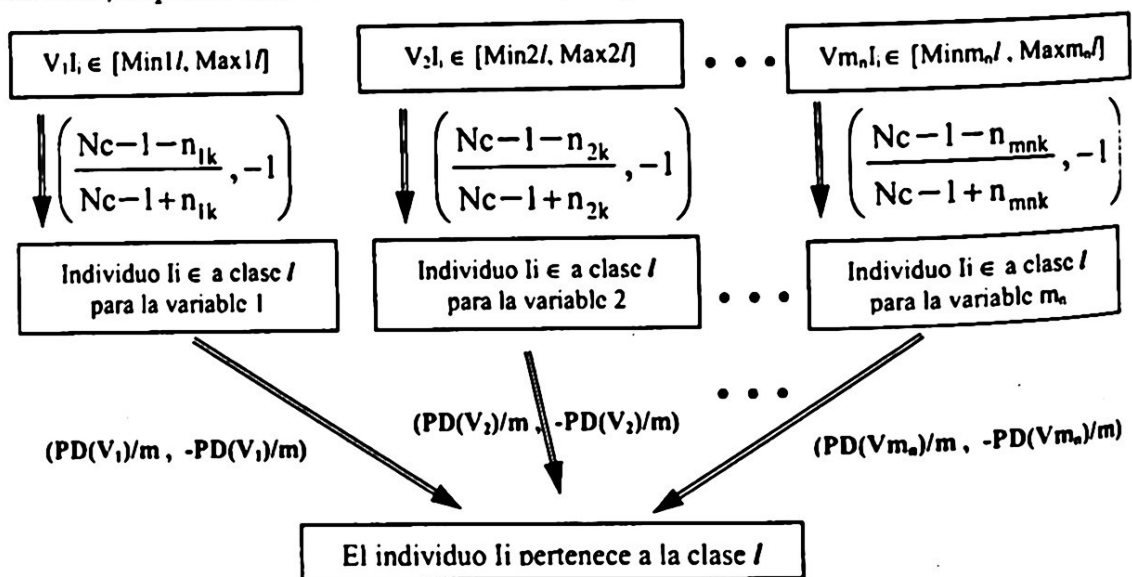


Figura 2.- Estructura de reglas para el Método de Descomposición de la Información

Si denotamos como $PD(V_j) \in [0,1]$ al peso de la variable V_j entonces la Regla Tipo 2 se puede expresar de la siguiente forma: $PTipo2'_{ij} \Rightarrow PTipo3'_i (PD(V_j)/m, -PD(V_j)/m)$;

Si analizamos las estructuras que se obtienen para las proposiciones y reglas que se han generado hasta este punto, se obtiene un esquema como el representado en la figura 2, donde se observa la estructura obtenida para un objetivo.

5.1.4. Análisis para Variables No Numéricas

Definición 4.- Una variable no numérica j (cualitativa o binaria) para el individuo I_i con valor V_{ij} pertenece a la clase l si $V_{ij} = X_{kj}$, donde X_{kj} es el valor del individuo k para la variable j y $k \in l$.

Como se observa, el criterio planteado considera que si se encuentra algún individuo en clase l tal que su valor para la variable j coincide con el valor correspondiente de la misma variable para I_i , entonces esto constituirá un elemento favorable con relación a la clasificación en esta clase.

Se puede entonces construir una nueva proposición de Tipo4 como sigue:

Proposición general Tipo4: La variable no numérica j para el individuo I_i pertenece a clase l . Se denota $PTipo4'_{ji}$.

Y a partir de ella definir la siguiente regla para la contribución a la proposición Tipo2:

Regla Tipo 3 (RT3): $PTipo4'_{ji} \Rightarrow PTipo2'_{ij} (W_l, -1)$;

Existen varias formas de definir la contribución W_l en este tipo de regla para el caso de variables no numéricas; dos posibilidades son las siguientes:

Caso 1.- $W_l = 1$. Se considera que al existir un individuo en la clase con valor igual al de I_i para j , esto asegura la pertenencia de I_i para la variable j a la clase l .

Caso 2.- $W_l = \frac{NE_{lji}}{NE_l}$ donde NE_{lji} denota la cantidad de individuos de la clase l cuyos

valores para la variable j son iguales al del individuo a clasificar I_i (V_{ij}) mientras que NE_l denota la cantidad de individuos de la clase l . En este caso se considera como importante la cantidad de individuos con este valor para la variable j .

5.2. Criterios de los Individuos más Semejantes

El método de los k-vecinos más cercanos se encuentra entre los más conocidos y usados Reconocimiento Supervisado de Patrones [6], [13]. En nuestro caso hemos considerado varios criterios que permiten incorporar a la estructura anterior las ideas de este método.

Consideraremos el conjunto $I_{knn} = \{X_1, X_2, \dots, X_{knn}\}$ de los knn patrones más semejantes a I_i (Individuo a clasificar) dentro de la Matriz de Aprendizaje M ($1 \leq knn \leq n$, un parámetro suministrado por el usuario y $X_k \in M$, $k=1, \dots, knn$) y supondremos que este conjunto está ordenado de acuerdo a su semejanza con I_i de mayor a menor, es decir, se parece más a I_i que X_2 y X_2 más que X_3 y así sucesivamente.

Criterio 1.- Si $X_k \in I_{knn}$, $X_k \in l$ y $SX_k I_i \in [0,1]$ denota la semejanza o parecido entre X_k e I_i entonces, esto contribuye en $SX_k I_i$ a la pertenencia de I_i a la clase l .

partir de este criterio se puede construir una regla considerando las proposiciones:

PTipo5: El individuo $X_k \in I_{knn}$.

PTipo6: El individuo $X_k \in l$.

Regla Tipo 4 (RT4): $PTipo5 \& PTipo6 \Rightarrow PTipo3'_i (SX_k I_i, 0)$;

Esta regla se interpreta como sigue: Si se cumple que el individuo X_k perteneciente a M_i se encuentra entre los knn más semejantes a I_i y X_k es miembro de la clase l , entonces esto contribuye con el valor de la semejanza $SX_k I_i$, a la pertenencia de I_i a la clase l . Se puede observar, que no todas las clases recibirán contribuciones producto de esta regla, dado que puede ocurrir que ninguno de sus elementos este en el conjunto I_{knn} , lo cual pudiera ser considerado como evidencia negativa en lugar de 0.

Criterio 2.- Si $X_k \in I_{knn}$ y $X_k \in l$ entonces esto contribuye en $CertI - CertI(k-1)/knn$ a la pertenencia de I_i a la clase l , donde $CertI$ es un parámetro externo que denota un valor de certidumbre inicial que permite controlar la influencia de este criterio.

Ejemplo: Si suponemos que $knn = 3$ y $CertI = 0.8$ entonces X_1 contribuye en 0.8, X_2 contribuye en 0.53 y X_3 contribuye en 0.27 lo cual implica una contribución menor según exista un menor parecido con I_i .

De la misma forma que en el caso anterior, se construyen proposiciones y una regla general para este criterio (RT5).

Criterio 3.- El usuario puede suministrar como datos las contribuciones de cada uno de los integrantes del conjunto I_{knn} (RT6).

Criterio 4.- Utilizar los mismos criterios anteriores, pero obteniendo conjuntos de individuos más semejantes a cada clase en lugar de a la matriz de aprendizaje completa. En este caso se tendrían reglas (RT7) para cada clase en particular.

5.3. Criterios del Análisis Global por Grupos

La Regla del Análisis Global por Grupos surge como una variante a la regla de clasificación clásica de los k vecinos más cercanos, en la cual se considera el cálculo de semejanzas con toda la clase en general, en lugar de obtener semejanzas individuales [8].

La semejanza entre un individuo I_i a clasificar y una clase l de la Información de

Aprendizaje, según esta regla, se obtiene de la siguiente forma: $S_{I_i l} = \sum_{j=1}^m 1 - \frac{1}{1 + dI_i^j l_j}$

donde $dI_i^j l_j$ representa una semejanza entre el valor de la variable j de I_i y los valores de esta variable, para todos los integrantes de la clase l , para lo cual también se proponen distintas variantes [8].

Lo interesante de esta regla de clasificación consiste, en que se demuestra que la expresión de semejanza entre un individuo y una clase ($S_{I_i l}$) cumple con los axiomas de una distancia clásica.

Desde el punto de vista del presente trabajo la idea consiste en permitir la integración de este procedimiento a la estructura anterior, lo cual se puede ejecutar de varias maneras. Algunos criterios son los siguientes:

Criterio 1.- La semejanza entre I_i y una clase l contribuye en $1/(1 + S_{I_i l})$ a la pertenencia de I_i a la clase l . La regla correspondiente a este criterio se expresa como: $0 \Rightarrow PTipo3^j_i(1/(1 + S_{I_i l}), 0)$; lo cual indica que bajo cualquier condición existe una contribución de $1/(1 + S_{I_i l})$ a la pertenencia de I_i a la clase l .

Criterio 2.- Se puede definir que solamente las clases de mayor semejanza reciban contribuciones positivas, mientras que el resto reciban contribuciones negativas.

Esto es, si la clase l se encuentra entre las k/l de mayor semejanza al individuo I_i (k/l es un parámetro externo) entonces existirá una contribución favorable a la clasificación proporcional a la semejanza y en caso contrario una contribución negativa.

Si definimos una proposición como sigue:

PTipo7: La clase l se encuentra entre las k/l de mayor semejanza al individuo I_i .

Se obtiene la Regla Tipo 8 (RT8): $PTipo7 \Rightarrow PTipo3'_i (1/(1+S_{I_i}l), 1 - 1/(1+S_{I_i}l))$;

Existen otras variantes definidas en el algoritmo original donde se utiliza para el cálculo las semejanzas individuo-classes, solamente aquellos integrantes de la clase que sean suficientemente parecidos de forma individual al patrón que se desea clasificar. Esta variante se denomina Análisis Global Distinguido [8] y también se pueden incorporar a las estructuras anteriores utilizando las mismas ideas presentadas.

6. Clasificación Colaborativa

Como se puede deducir del estudio de los epígrafes anteriores, la idea que se presenta consiste en construir una BC constituida por un conjunto de Proposiciones y un conjunto de Reglas de Producción, donde estas últimas son la representación de diversos métodos de clasificación supervisada.

El usuario del sistema que implementa esta teoría, tendrá la posibilidad de escoger los criterios o métodos a utilizar, que pueden ser uno o varios, de forma tal que el resultado de clasificación será la combinación de las contribuciones que aporta cada uno de los métodos a los objetivos finales, obteniéndose como resultado la seguridad que se tiene en relación con la pertenencia del patrón a clasificar a cada clase de la Información de Aprendizaje.

Para ejecutar el procesamiento de los grados de certidumbre se ha tomado una teoría utilizada para el desarrollo de Sistemas Expertos, que generaliza otras teorías clásicas en este campo, mediante una aproximación algebraica [10], [11].

Si se considera que $x_1, x_2, \dots, x_r$ denotan valores de certidumbre provenientes de las contribuciones de r reglas de producción, entonces se necesita definir una operación $\oplus_o$ (o denota el tipo de operación) que permita combinar estos valores dos a dos y que cumpla con las siguientes propiedades:

1. $x_i \oplus_o x_j \in [-1, 1]$ (operación interna)
2. Asociatividad: $x_i \oplus_o (x_j \oplus_o x_k) = (x_i \oplus_o x_j) \oplus_o x_k$
3. Conmutatividad: $x_i \oplus_o x_j = x_j \oplus_o x_i$
4. Existencia de elemento neutro: $x_i \oplus_o 0 = x_i$
5. Existencia del opuesto: $x_i \oplus_o -x_i = 0$
6. Monotonía: Si $x_i \leq x_j$ entonces $x_i \oplus_o x_k \leq x_j \oplus_o x_k$
7. $x_i \oplus_o 1 = 1$, $x_i \neq -1$; $x_i \oplus_o -1 = -1$, $x_i \neq 1$
8. $-1 \oplus_o 1$ indefinido

Como se puede deducir, esta operación sobre el conjunto de definición de valores de certidumbre $[-1, 1]$, define una estructura algebraica que es un Grupo Abeliano.

Para obtener una operación como ésta se busca un isomorfismo f , con el grupo de los números reales con la operación suma de la siguiente forma:

$$x_i \oplus_o x_j = f^{-1}(f(x_i) + f(x_j)) \quad (f^{-1} \text{ denota la inversa de } f)$$

Existen diversas funciones que pueden ser usadas en esta operación; dos ejemplos interesantes para los que se demuestra que las operaciones resultantes son Grupos

Abelianos isomorfos a los utilizados en los Sistemas Expertos clásicos Prospector Mycin [20] son: $f_1(x) = \text{Ln}[1/(1-x)]$ y $f_2(x) = \text{Ln}[(1+x)/(1-x)]$ respectivamente.

Ejemplo: Si se obtienen contribuciones de -0.2 y 0.6 producto de la aplicación de dos reglas, entonces la contribución global de ambas utilizando f_1 es:
 $-0.2 \oplus 0.6 = f^{-1}(f(-0.2) + f(0.6)) = 0.455$.

7. Incorporación de Conocimientos

Con la teoría propuesta, además de obtenerse un procedimiento para clasificación colaborativa entre métodos de Reconocimiento Supervisado, con la teoría propuesta, también se logra una estructura semejante a la de una BC construida utilizando experiencias humanas.

Lo anterior tiene diversas implicaciones, la primera tiene que ver con el proceso clasificación, puesto que no existe ningún problema en agregar los criterios de expertos humanos, mediante reglas de producción.

Supongamos que se agrega una regla de producción suministrada por los expertos como la siguiente:

$\text{Ant} \Rightarrow \text{PTipo3}_i(W_1, W_2)$; donde "Ant" representa una proposición o combinación de proposiciones y W_1, W_2 son los pesos o grados de certidumbres para la afirmación negación absoluta de Ant, respectivamente.

Si consideramos que W_p denota la contribución obtenida producto de la aplicación uno o varios criterios de reconocimiento y W_e la contribución obtenida después de aplicar una o varias reglas suministradas por los expertos humanos, entonces el resultado combinar ambas contribuciones es: $W_p \oplus W_e$, lo cual brinda un resultado híbrido.

7.1. Elaboración de los Resultados

Otro aspecto a considerar es la respuesta final que debe brindar el sistema diseñado esta forma, puesto que como ya sabemos, lo que se obtiene es la seguridad, favorable desfavorable de la pertenencia de un patrón a clasificar a cada una de las clases existentes en la Información de Aprendizaje.

Sin embargo, la solución a esta situación no es única, ya que guarda cierta relación el problema particular que se este tratando.

Existen algunos criterios de conclusión suministrados por el sistema que el usuario puede elegir. Como ejemplos tenemos los siguientes:

Criterio 1.- Brindar como clasificación aquella clase con mayor valor de certidumbre positiva.

Criterio 2.- Brindar como solución hasta un máximo de C_r clases cuyos valores certidumbre sean máximos valores positivos y la diferencia entre la de mayor valor la de menor no exceda a W_d (valor externo).

Criterio 3.- Presentar la clasificación como dudosa si no existe una clase cuya diferencia con las demás es significativa. El concepto de significativo lo define el usuario introduciendo un parámetro de diferencia W_d entre la clase de mayor peso y la siguiente en valor.

Así por ejemplo, si se está ante una situación donde no se admite la pertenencia de patrón a más de una clase, entonces el primer criterio puede resultar apropiado y se puede

introducir una variante de clasificación dudosa, para el caso en que dos o más clases estén muy cerca en los valores de certidumbre, lo cual es una combinación con el tercer criterio.

Finalmente, puede existir también el caso para el cual ninguno de los criterios resulte apropiado, ante tal situación el sistema propuesto permite que el usuario defina su proceso de elaboración de las conclusiones, utilizando una teoría que está diseñada especialmente con ese objetivo y que es parte de un ambiente para el desarrollo de sistemas inteligentes, denominado HAries [9], [10], [11].

Esto es posible ya que el sistema de Reconocimiento Supervisado que se propone puede trabajar de forma independiente o de forma colaborada con HAries, lo cual es necesario si se incorpora conocimiento humano al sistema. Esto ha permitido aprovechar las herramientas propias de los sistemas basados en conocimientos para construir sistemas híbridos.

8. Consideraciones Finales

Para lograr la combinación de varios clasificadores se ha construido una teoría que propone la construcción de una Base de Conocimientos a partir de la Información de Aprendizaje, para métodos de Reconocimiento Supervisado de Patrones.

Se observa, que no se realiza una combinación de los resultados provenientes de diversos métodos, sino que se utilizan sus teorías para producir contribuciones a la clasificación, cuyo resultado final es la combinación global de todas.

Se puede asegurar, que la arquitectura generada constituye una plataforma de trabajo en la que se pueden integrar otros métodos o criterios de clasificación supervisada, lo cual posibilita la ampliación de sus alcances y un ambiente para la experimentación.

Un caso típico son los árboles de decisión [20], puesto que a partir de estos se puede obtener reglas de producción con la misma estructura que las explicadas en los epígrafes anteriores.

Otro ejemplo son los Algoritmos de Votación o de Zhuravliov [17] cuya idea fundamental yace en considerar que un objeto puede parecerse a otro, pero no en su totalidad y esto puede brindar información para la clasificación buscada.

De hecho en estos algoritmos ya se considera, como su nombre lo indica, ciertas ideas de clasificación colaborada y resulta muy natural también su incorporación a la arquitectura que se propone.

Estos resultados son muy útiles desde el punto de vista práctico, ya que posibilitan la inclusión de conocimientos humanos como parte del proceso de clasificación o a la inversa, de manera que se pueda utilizar una Base de Datos como parte del diseño de un Sistema Experto, de forma tal que los resultados sean producto de la combinación de ambas fuentes de información.

No obstante, quedan todavía diversas preguntas planteadas, como por ejemplos: ¿La incorporación de muchos métodos mejora la efectividad de la clasificación?, los resultados obtenidos en diversos trabajos apuntan a que la unificación tiende a mejorar los resultados, pero, ¿cuáles métodos usar ante un problema dado y hasta qué punto incorporar más métodos provoca una mejora en los resultados? sigue siendo aún, un problema abierto.

Referencias

- [1] Alexandre, L., Aurelio C. Campilho, A., Kamel, M.: "On combining classifiers using sum and product rules". *Pattern Recognition Letters*, Vol. 22 (12) pp. 1283 – 1289 (2001)
- [2] Aliand, K. M., Pazzani, J. "On the link between error correlation and error reduction in decision tree ensembles". Technical Report UCI TR-95-38, University of California, Irvine, California (1995)
- [3] Bauer, E., Kohavi, R.: "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants". *Machine Learning*, 36(1/2): 105-139 (1999)
- [4] Breiman, L.: "Bagging Predictors". *Machine Learning*, 24(2): 123-140 (1996)
- [5] Cho, S. B., Jin Kim, J.: "Combining multiple neural networks by fuzzy integral for robust classification". *IEEE Transactions on Systems, Man and Cybernetics*, 25(2): 380-384 (1995)
- [6] Cover, T.M., Hart, P.E.: "Nearest Neighbor Pattern classification". *IEEE trans. Information Theory*, IT-13 Pp 21-27 (1967)
- [7] De la Cruz, A. V.: "Un sistema de algoritmos para reconocimiento de patrones supervisado con posibilidades de auto-aprendizaje. Ejemplos de aplicación". IV Jornada Científica del Instituto de Geofísica y Astronomía A.C.C. 18-20 Octubre (1984)
- [8] De la Cruz, A. V.: "SRPS: Un sistema para reconocimiento de patrones supervisado" *Informática'90*, Reporte de participación. Pp 227-240 (1990)
- [9] De la Cruz A.V., Valdés J.J., Pérez A., Jócik E., Balsa J., Rodríguez A., "The ARIES Environment for the Development of Knowledge Based Expert Systems". *Industrial Applications of Artificial Intelligence*. Elsevier Science Publishers. B.V. North-Holland, pp 203-209 (1991)
- [10] De la Cruz, A.V., Valdés J.J., Jócik E., Balsa J., Rodríguez A.: *Fundamentos y Práctica de la Construcción de Sistemas Expertos*. Editorial Academia, La Habana, Cuba (1993)
- [11] De la Cruz, A.V.: *Construcción Práctica de Sistemas Expertos*. Memorias Foro Computación de la teoría a la práctica, México DF, pp. 262-278 (1999)
- [12] Dhillon, I.S., Modha, D.S.: "A Data-Clustering Algorithm on Distributed Memory Multiprocessors". *Large-Scale Parallel Data Mining. Lecture Notes in Artificial Intelligence 1759*. Subseries of Lecture Notes in Computer Science. Pp 245-260 (2000)
- [13] Duda, R. O., Hart, P. E.: "Pattern Classification and Scene Analysis". John Wiley, New York (1973)
- [14] Freund, Y., Schapire, R.: "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". *J. Computer and System Sciences*, 55:119-139 (1997)
- [15] Johnson, E.L., Kargupta, H.: "Collective, Hierarchical Clustering from Distributed, Heterogeneous Data". *Large-Scale Parallel Data Mining. Lecture Notes in Artificial Intelligence 1759*. Subseries of Lecture Notes in Computer Science. Pp 221-244 (2000)
- [16] Krogh, A., Vedelsby, J.: "Neural network ensembles cross validation, and active learning" *NIPS 7*. Lee 97 Peter M. Lee 1997. Bayesian Statistics. New York: Wiley (1995)
- [17] Ruíz, J., Guzmán, A., Martínez, J. F.: "Enfoque lógico combinatorio al reconocimiento patrones". Instituto Politécnico Nacional, México (1999)
- [18] Schaal, S., Christopher G.: "From isolation to cooperation: an alternative view of a system experts". *NIPS 7* (1995)
- [19] Schalkoff, R.: "Pattern Recognition Statistical, Structural and Neural Approaches". John Wiley and Sons, Inc. 364 Pp. (1992)
- [20] Shapiro S. C.: "Encyclopedia of Artificial Intelligence". Volume 1. John Wiley & Sons (1992)
- [21] Tumer, K., Joseph Ghosh, J.: "A framework for estimating performance improvements hybrid pattern classifiers". *Proceedings of the World Congress on Neural Networks*, III:220-225 (1994)
- [22] Valev, V., Asaithambi, A.: "Multidimensional pattern recognition problems and combining classifiers". *Pattern Recognition Letters*, Vol. 22 (12) pp. 1291 – 1297 (2001)