UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS

TESIS DOCTORAL

Relevancia y Calidad en Sistemas de Clasificación Borrosa. Un Enfoque Algebraico

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Fabián Alberto Castiblanco Ruiz

DIRECTORES

Francisco Javier Montero de Juan
Juan Tinguaro Rodríguez González
Camilo Andrés Franco de los Ríos

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS

TESIS DOCTORAL

Relevancia y Calidad en Sistemas de Clasificación Borrosa. Un Enfoque Algebraico

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

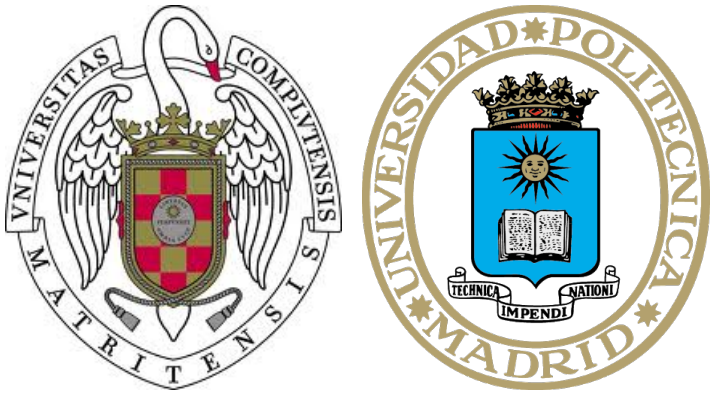
PRESENTADA POR

Fabián Alberto Castiblanco Ruiz

DIRECTOR

Francisco Javier Montero de Juan
Juan Tinguaro Rodríguez González
Camilo Andrés Franco de los Rios

Programa de Doctorado en Ingeniería Matemática, Estadística e Investigación operativa por la Universidad Complutense de Madrid y la Universidad Politécnica de Madrid.



Relevancia y Calidad en Sistemas de Clasificación Borrosa. Un Enfoque Algebraico

Tesis Doctoral

FABIÁN ALBERTO CASTIBLANCO RUIZ

Directores:

FRANCISCO JAVIER MONTERO DE JUAN
JUAN TINGUARO RODRÍGUEZ GONZÁLEZ
CAMILO ANDRÉS FRANCO DE LOS RIOS

Año 2020

Agradecimientos

Deseo ofrecer mis más sinceros agradecimientos al profesor Javier Montero porque me brindó la oportunidad de trabajar con él y aprender de sus palabras siempre sabias. Su apoyo y confianza en mi trabajo y su capacidad para orientar y guiar las ideas han sido un aporte invaluable, tanto en lo académico como en lo personal. En medio de su inconmensurable sapiencia, siempre brilló la humildad, la palabra precisa, la mano amiga, la pregunta inquietante y una voz de aliento. Siempre será un ejemplo a seguir.

Agradezco de forma sincera al profesor Tinguaro Rodríguez por brindarme su apoyo y dedicación. Cada parte del proceso fue un instante de crecimiento gracias a sus valiosos aportes. Su rigurosidad, sabiduría y vigilancia ante los detalles hicieron del trabajo investigativo todo un proceso de formación. Agradezco su paciencia y rigurosidad ante cada paso que he logrado dar. La consolidación de las ideas no habrían sido posibles sin su entrega y admirable contribución en cada línea de la tesis. Así se reconoce un buen maestro.

Deseo expresar de manera especial y sincera mis agradecimientos al profesor Camilo Franco, por su actitud siempre dispuesta y comprensiva a mi proceso de formación y al desarrollo de la investigación. Los encuentros siempre enriquecedores y abundante en ideas han hecho posible la consolidación de la tesis. Sus observaciones siempre oportunas, precisas y colmadas de conocimiento hicieron viable el desarrollo de un trabajo gratificante. Espero poder seguir aprovechando todo su potencial.

Finalmente, deseo agradecer a la Universidad Complutense de Madrid y a la facultad de ciencias matemáticas por brindarme la oportunidad de ser un integrante de tan prestigiosa comunidad. A la doctora Begoña Vitoriano, por su apoyo y colaboración en cada parte del proceso.

Dedicatoria

Dedico el fruto del trabajo investigativo a mi familia, quienes han sido mi soporte y mi bastión en cada etapa de mi vida. A mi madre, por ser quien me da fortaleza y nunca me deja desfallecer. Es mi motor de vida. A mi padre, porque me impulsa a ser cada día mejor y a ver la vida como única. A mi hermana, porque con su amor ha colmado mi vida de alegría.

Dedico especialmente este trabajo a Yuly Franco, mi compañera de vida, porque no ha sido fácil convivir con un doctorando, se ha sacrificado tiempo y atención pero sobretodo, ha sido su comprensión y amor lo que me ha permitido seguir adelante. Has estado en mis derrotas y es menester dedicarte este logro.

A mis amigos y compañeros de vida, a Juan Diego, a Luis Eduardo, a Yesid y a todos aquellos que me acompañaron en el camino. Sé que no ha sido fácil lidiar con mis pasiones cuando hablaba de lo borroso.

Índice general

Índice de figuras

Índice de Tablas

Resumen en castellano

La validación de clústeres borrosos es un problema abierto y con varias perspectivas para su estudio. La presente tesis doctoral aborda el problema de determinar la calidad de particiones borrosas en procesos de clasificación no supervisada. Para tal objetivo, se han propuesto criterios e índices que permiten determinar particiones borrosas de calidad y de alta calidad, la relevancia de una clase o familia de clases en una partición borrosa dada y, el número óptimo de clases de una partición de acuerdo a algunas características estudiadas.

La propuesta establecida, emplea herramientas provenientes del álgebra abstracta como son, la teoría de grupos, particularmente la teoría de subgrupos borrosos, matrices de relaciones y los polinomios característicos de tales matrices. Los resultados teóricos fueron aplicados y validados en el problema de segmentación de imágenes.

Los índices y criterios propuestos presentan un enfoque alternativo a las teorías existentes sobre la evaluación de particiones borrosas y, los resultados hallados muestran algunas ventajas en contraste con el uso de un conjuntos de índices propuestos en la literatura.

Abstract

Fuzzy cluster validation is an open problem with several perspectives for study. This doctoral thesis addresses the problem of determining the quality of fuzzy partitions in unsupervised classification processes. For this purpose, criteria and indexes have been proposed that allow determining quality and high-quality fuzzy partitions, the relevance of a class or family of classes in a given fuzzy partition and, the optimal number of classes of a partition according to some characteristics studied.

The proposal uses tools from abstract algebra such as group theory, particularly the theory of fuzzy subgroups, matrices of relations and the characteristic polynomials of such matrices. The theoretical results were applied and validated in the image segmentation problem.

The proposed indices and criteria present an alternative approach to the existing theories on the evaluation of fuzzy partitions and, the results found show some advantages in contrast to the use of a set of existing indices found in the literature.

Introducción

Desde los inicios de la teoría de la lógica borrosa propuesta por Zadeh, [118] los sistemas de clasificación han presentado grandes avances en su desarrollo teórico y práctico; ejemplo claro de ello han sido los trabajos presentados en [5, 9, 15, 33, 54, 92]. El acercamiento desde el enfoque borroso en estas áreas, ha permitido la comprensión de información manejada por los usuarios en entornos donde las fronteras o límites entre subconjuntos no están claramente definidos, y por ende, ha sentado las bases para el tratamiento de información cuya naturaleza es borrosa y requiere de grados de verificación. En particular, los problemas de teledetección o segmentación de imágenes han logrado grandes avances mediante el desarrollo de sistemas de clasificación borroso como los presentados en [4, 5, 18].

El sistema de clasificación borroso desarrollado por A. del Amo, Montero, Biging y Cutello, [5] (ver también [4]), parte de una generalización del concepto de partición borrosa propuesto por Ruspini [100] y define dicho sistema de clasificación mediante familias de funciones de agregación actuando sobre los grupos o clústeres de una partición borrosa. Tales grupos conforman lo que se denomina una familia de clases borrosas.

El enfoque de sistemas de clasificación borrosa, propone índices que permiten medir la redundancia, relevancia y cubrimiento de las clases obtenidas en una partición; es decir, se construye un sistema que a su vez permite en una primera etapa, evaluar la clasificación establecida.

Sin embargo, la evaluación de la calidad de una partición es un problema que aún requiere ser abordado con mayor profundidad (ver [78]) y, con especial énfasis al considerar los sistemas de clasificación borrosa y su estudio sobre cada uno de los índices propuestos. En particular, se hace relevante y pertinente un estudio más amplio sobre la propiedad de la *relevancia* de la familia de clases establecidas mediante

tal partición borrosa; un campo en desarrollo y aún por explorar.

Específicamente, el anterior planteamiento pone de manifiesto uno de los propósitos de la presente tesis; un estudio profundo sobre los sistemas de clasificación borrosa y su potencial para generar o encontrar nuevas estructuras matemáticas o estructuras heredadas, que potencien al mismo tiempo el aprendizaje sobre los procesos de clasificación de objetos, así como las propiedades que poseen las particiones obtenidas bajo dichos procesos.

En el marco del aprendizaje no-supervisado, tanto el reconocimiento de patrones como la agrupación de elementos u objetos, son áreas del conocimiento que gracias a los desarrollos tecnológicos han presentado una evolución acelerada, en particular, en las labores enfocadas al procesamiento de imágenes considerando el enfoque borroso (ver [106]).

Muchas metodologías de procesamiento de imágenes están diseñadas para identificar una partición de todos los píxeles de la imagen de acuerdo con la intensidad de los mismos. Sin embargo, la forma como un experto determina una partición e identifica los grupos de dicha partición, permite una variedad de posibilidades; si bien el trasfondo matemático de una clasificación es un punto de verificación esencial, el contraste con los resultados obtenidos por un experto puede implicar variaciones, más aún en el ámbito de lo borroso.

Una clase o clúster dentro de una partición en imágenes, corresponde a un conjunto de píxeles similares, es decir, un conjunto de píxeles que comparten alguna propiedad (propiedad que puede ser relativa a otros píxeles, como sucede cuando se buscan bordes). Por ejemplo, en ocasiones la similitud de los píxeles se establecen mediante algún tipo de patrón predefinido, pero a veces estos patrones se establecen directamente sobre la imagen de acuerdo con ciertos criterios. Adicionalmente, se pueden imponer ciertas restricciones de forma, por ejemplo, que los píxeles dentro de la misma clase estén conectados para definir una región lo suficientemente homogénea, que debería

estar lo suficientemente diferenciada de otras regiones. Esto podría imponerse debido a un conocimiento previo sobre esos patrones, o porque esas formas son de alguna manera convenientes para los intereses o limitaciones del decisor. En la misma línea, el responsable de la toma de decisiones puede tener dificultades para gestionar demasiadas categorías o regiones, y podría darse el caso que el resultado declarado por el responsable de la decisión requiera alta precisión o confiabilidad en ciertos aspectos. Si aparecen restricciones de forma, se debe tener en cuenta la posición relativa de los píxeles.

En cualquier caso, son necesarias diversas medidas que permitan evaluar en qué grado una posible clase es una buena candidata para ser declarada como clase o clúster, y en qué medida una familia de posibles clases satisface los intereses y limitaciones de los tomadores de decisiones, incluyendo hasta qué punto los objetos se *explican* a partir de esas clases y en qué medida se diferencian las clases. En otras palabras, se deben evaluar tanto las propiedades locales de cada clase como las propiedades globales de todo el sistema.

De acuerdo con lo anterior, se pone en evidencia la complejidad de procesar una sola imagen, aún sin introducir movimiento o secuencias de imágenes. No es de esperarse entonces que la calidad de la partición generada sobre una imagen pueda evaluarse adecuadamente mediante una única medición o un único índice. Un problema con múltiples facetas necesita múltiples medidas. Aunque sea natural elegir un solo criterio y buscar la optimización de la medida asociada, tal simplificación podría estar dejando de lado la visión compleja del tomador de decisiones, quien sobre todo suele reclamar información útil y confiable. Además, es cierto que en algunos contextos se buscan, o son deseables, decisiones impulsadas automáticamente, pero incluso estas decisiones automáticas generalmente se basan en una batería de índices específicos, por lo que todo el proceso puede ser monitorizado por un ser humano.

Lo expuesto en la líneas anteriores, anidado con los sistemas de clasificación borro-

sa, pone en contexto el segundo propósito general de la presente tesis; la construcción de índices y procesos que permitan evaluar una partición borrosa considerando y profundizando en el uso de los conceptos de cubrimiento, redundancia y relevancia, aplicados en el procesamiento de imágenes.

Si bien, la evaluación de la calidad de una partición borrosa requiere de distintos enfoques y herramientas para su determinación, y aunque existe una amplia variedad de índices (ver [16]), se hace necesario y pertinente establecer mecanismos que integren más características de una partición en una misma secuencia o índice. En este sentido, se propone extender las estructuras dadas por los sistemas de clasificación borrosa para evaluar la calidad de una partición.

Una exploración sobre la constitución de las estructuras matemáticas propias y subyacentes a los sistemas de clasificación borrosa permite, en general, descubrir nuevos horizontes en investigación y riqueza en propiedades concebidas desde su estudio y desde su aplicabilidad. En este sentido, las herramientas proporcionadas por el álgebra se presentan como una alternativa adecuada para abordar e identificar propiedades de las estructuras matemáticas específicas y determinar relaciones entre sus elementos.

En particular, la propuesta de la tesis gira alrededor de la identificación de una estructura de grupo conmutativo sobre los sistemas de clasificación borrosa, la cual es dotada de una noción de medida en términos de la similitud y disimilitud de sus elementos. En consecuencia, se indaga por las relaciones, propiedades y resultados en términos algebraicos que se pueden encontrar y, la posibilidad de extrapolación a la evaluación de particiones borrosas.

De acuerdo con lo anterior, la presente tesis se encuentra estructurada en los siguientes capítulos. El capítulo 1 denominado *Preliminares*, establece tanto el estado del arte sobre el tema central de la clasificación borrosa como los fundamentos matemáticos que permiten consolidar la propuesta desarrollada en esta tesis. En la Sección 1.1, se presentan las generalidades sobre la teoría de los *operadores de agregación* debi-

do a su directa relación con las estructura de los sistemas de clasificación. Se enfatiza únicamente en los elementos necesarios para el desarrollo de la tesis. Posteriormente, se aborda lo concerniente a las *medidas de similitud y disimilitud* estableciendo una clasificación de acuerdo a sus propiedades. Tales medidas permiten establecer dos enfoques o caminos diferenciados sobre la propuesta presentada.

En la Sección 1.3 del Capítulo 1, se presentan los elementos necesarios sobre la teoría de los *subgrupos borrosos*, y se exponen las definiciones y resultados que permiten integrar tal teoría con los procesos de clasificación. En la Sección 1.4, se profundiza en la revisión literaria sobre el área de la clasificación borrosa, haciendo especial énfasis en el enfoque de aprendizaje no supervisados, lo cual corresponde con el marco general de la propuesta. Por lo tanto, se abordan específicamente los algoritmos de agrupamiento iterativo y algunas medidas existentes para su evaluación. Finalmente, en la Sección 1.5 se expone la herramienta fundamental y punto de partida para la propuesta, los *sistemas de clasificación borrosa* propuestos en [5], presentando los resultados alcanzados hasta la fecha por los autores.

A partir del Capítulo 2, denominado *Similitud y disimilitud sobre un grupo conmutativo de operadores de agregación*, se establecen los primeros resultados de la tesis. A partir de este punto, lo expuesto corresponde a los hallazgos y elementos de construcción propia producto de la investigación reflejada en esta tesis. En la Sección 2.1, se propone una definición sobre lo que se considera el *grado global* de propiedades evaluadas sobre una partición borrosa. Así, se establecen por ejemplo los grados globales de cubrimiento y los grados globales de solapamiento de una partición. En la Sección 2.2, se construyen dos operadores de agregación, en particular operadores recursivos, sobre los sistemas de clasificación borrosa. Tales operadores permiten definir una operación entre ellos y por lo tanto, una estructura de grupo conmutativo formada por operadores. Dicha estructura conforma el eje central de los desarrollos subsiguientes. En la Sección 2.3, se dota al grupo conmutativo de una relación de si-

militud y se halla una relación directa con los subgrupos normales borrosos formados sobre el grupo. De igual forma, para finalizar el Capítulo 2, se dota de una función de disimilitud al grupo conmutativo, estableciendo matrices de disimilitud y estudiando algunas propiedades de los polinomios característicos para tales matrices.

A partir de los resultados del capítulo anterior, en el Capítulo 3, denominado *Criterios de relevancia y calidad para particiones borrosas*, se proponen dos perspectivas para evaluar particiones borrosas. Por un lado, considerando relaciones de similitud se propone un algoritmo que permite determinar cuándo una partición es de *calidad* y cuándo sus clases son relevantes. Por otro lado, a partir de las relaciones de disimilitud, se establece un índice que permite determinar el número *óptimo* de clases que debe tener una partición borrosa. Este índice se obtiene como resultado de un proceso de optimización.

Finalmente, en el Capítulo 4, se aplican las propuestas del Capítulo 3 sobre un problema de segmentación de imágenes con el fin de validar los resultados y contrastarlos con algunos índices existentes. Se emplean tanto imágenes a color como tomografías computarizadas y se interpretan los resultados de la aplicación.

Objetivos

Objetivo general. Construir criterios e índices que permitan evaluar la calidad de una partición borrosa a partir de las propiedades proporcionados por un sistema de clasificación: relevancia, cubrimiento y redundancia.

Objetivos específicos

1. Determinar el estado actual de los sistemas de clasificación borrosa y de su uso en la evaluación de particiones borrosas.

2. Establecer un marco de referencia sobre el concepto de relevancia y su evaluación sobre las clases en una partición borrosa.

3. Establecer las estructuras y propiedades algebraicas anidadas o subyacentes a los sistemas de clasificación borrosa.

4. Relacionar las estructuras y propiedades con los procesos de validación de una partición borrosa.

5. Implementar y validar las técnicas propuestas sobre diversos problemas de segmentación de imágenes.

Capítulo 1

Preliminares

El presente capítulo establece los conceptos y resultados previos extraídos de la literatura que permiten desarrollar, en los capítulos posteriores, las ideas y resultados de la tesis. En las tres primeras secciones se enuncian resultados clásicos y necesarios de la lógica borrosa en general como lo son, operadores de agregación, las funciones de similitud, distancia y subgrupos borrosos. En las secciones posteriores del mismo capítulo, se describe el marco de referencia sobre el cual se establece la propuesta y se generan los aportes al conocimiento; la clasificación borrosa y los sistemas de clasificación borrosa.

1.1. Operadores de agregación

La teoría de los operadores de agregación es tan amplia como útil en diversos campos de las ciencias. El problema de agregar, fusionar o representar una cantidad diversa e incluso heterogénea de valores de entrada en un valor de salida, requiere de su delimitación para un abordaje más preciso. En particular, la presente sección está dedicada a esbozar algunas generalidades y nociones básicas sobre los operadores de

agregación con entradas finitas y sobre el conjunto de los números reales, específicamente, sobre el intervalo $[0, 1]$. Las definiciones aquí plasmadas corresponden a los elementos necesarios para el desarrollo de las ideas expuestas en la presente tesis a partir del Capítulo 2.

Definición 1. *[21] Un operador de agregación*

$$A : \bigcup_{n\in\mathbb{N}} [0,1]^n \to [0,1]$$

es una función que satisface:

1. $A\,(0,...,0) = 0$ *y* $A(1,...,1) = 1$,

2. $\forall n \in \mathbb{N}$, *si* $x_1 \leq y_1, ..., x_n \leq y_n$ *entonces,* $A\,(x_1,...,x_n) \leq A\,(y_1,...,y_n)$

De acuerdo con la Definición 1, para cada $n \in \mathbb{N}$ se define un operador de agregación n-ario A_n y, por lo tanto, A se entiende como una familia de operadores de agregación n-arios. Para simplificar, en lo que sigue se empleará el término operador de agregación para hacer referencia tanto a una familia de operadores de agregación n-arios como a los propios operadores n-arios.

Los operadores de agregación presentan diversas propiedades que permiten su clasificación y uso según las características que cumplen. En particular, considerando el trabajo desarrollado por Dombi [37, 38], los operadores con comportamiento conjuntivo y disyuntivo se establecen en tal sentido que el *principio de correspondencia* sea satisfecho, es decir, se dice que un operador de agregación A es conjuntivo si para cualquier conjunto finito de elementos $x_1, ..., x_n$ se cumple que $A(x_1, ..., x_n) = 0$ siempre que exista un $x_i = 0$. En la misma línea, se dice que el operador es disyuntivo si $A(x_1, ..., x_n) = 1$ siempre que exista un $x_i = 1$.

La anterior perspectiva conforma una versión más general que la planteada en [39] (incluso ver [45]) bajo la cual, una operador de agregación A presenta una acti-

tud conjuntiva si $A(x_1, ..., x_n) \leq \min\{x_1, ..., x_n\}$ y un operador presenta una actitud disyuntiva si $A(x_1, ..., x_n) \geq \max\{x_1, ..., x_n\}$.

En la misma línea, siguiendo con lo planteado en [21, 39] (ver también [45]), se definen algunas propiedades generales que pueden cumplir los operadores de agregación. A continuación se definen las propiedades mas relevantes para el desarrollo de la presente tesis.

Definición 2. *[39] Un operador A es llamado idempotente si satisface*

$$A(x, ..., x) = x, \text{ para todo } x \in [0, 1]$$

Definición 3. *[39] Un operador A es compensativo si y solo si*

$$\min\{x_1, ..., x_n\} \leq A(x, ..., x) \leq \max\{x_1, ..., x_n\}.$$

Definición 4. *[21] Un operador A es llamado simétrico si*

$$A(x_1, ..., x_n) = A(x_{\alpha(1)}, ..., x_{\alpha(n)})$$

para toda permutación $\alpha = (\alpha(1), ..., \alpha(n))$ de $1, ..., n$.

Definición 5. *[21] Un operador A es asociativo si $\forall n, m \in \mathbb{N}, \forall x_1, ..., x_n, y_1, ..., y_m$:*

$$A(x_1, ..., x_n, y_1, ..., y_m) = A(A(x_1, ..., x_n), A(y_1, ..., y_m))$$

Definición 6. *[21] Sea A un operador de agregación. Un elemento $e \in [0, 1]$ es*

llamado elemento neutro de A si, cuando $x_i = e$ para algún $i \in \{1, ..., n\}$, entonces

$$A(x_1, ..., x_n) = A(x_1, ..., x_{i-1}, x_{i+1}, ..., x_n)$$

A partir de las anteriores definiciones se establecen los operadores *promedio*:

Definición 7. *[45] Un operador de agregación A se denomina tipo promedio si A es un miembro de la clase de operadores compensativos y simétricos diferentes de los operadores* $\max\{x_1, ..., x_n\}$ *y* $\min\{x_1, ..., x_n\}$.

Definición 8. *[21] Sea A un operador de agregación. Entonces el operador de agregación denotado por A^d y dado por,*

$$A^d(x_1, ..., x_n) = 1 - A(1 - x_1, ..., 1 - x_n)$$

se denomina un operador dual de A.

En particular, si el operador A cumple que $A^d = A$, entonces el operador A se denomina *auto dual* o *suma simétrica*. Tales operadores se encuentran caracterizados por satisfacer la ecuación funcional

$$A(x_1, ..., x_n) + A(1 - x_1, ..., 1 - x_n) = 1.$$

Finalmente, se presenta un tipo de funciones particulares que hacen parte tanto del marco general de los sistemas de clasificación borrosa como de la propuesta de la tesis; las negaciones:

Definición 9. *[45] Una función $n : [0, 1] \to [0, 1]$ tal que $n(0) = 1$, $n(1) = 0$ y n es no creciente, se denomina una negación. Si n es estrictamente decreciente y continua, se*

denomina negación estricta. Adicionalmente, si n es una negación estricta y satisface que $n(n(x)) = x$, entonces se denomina una negación fuerte.

Definidas las negaciones, entonces se establece la siguiente definición denotando en particular a las negaciones fuertes por N.

Definición 10. *[45] Un operador A es estable por la negación fuerte N si*

$$A(N(x_1), ..., N(x_n)) = N(A(x_1, ..., x_n)).$$

De forma inmediata se observa que la Definición 10, es una generalización de la Definición 8 en tanto se establece para cualquier negación fuerte.

Establecidas las definiciones básicas sobre operadores de agregación, a continuación se presenta un marco de referencia general sobre las medidas de similitud y disimilitud que permite en particular, ubicar tales medidas en el contexto de lo borroso.

1.2. Medidas de similitud y disimilitud

Los enfoques de aprendizaje automático para el procesamiento de datos de forma supervisada o no supervisada, pueden basarse en la comparación de los objetos de análisis o sus características con respecto a su similitud o disimilitud. El resultado de tal comparación puede obtenerse ya sea por cálculos matemáticos o por algún tipo de juicio basado en las percepciones del evaluador o experto. Ejemplos de comparaciones definidas matemáticamente son el uso de medidas de distancia como la Euclidiana, la de Mahalanobis, la de Minkowski, medidas de correlación o divergencia, probabilidades conjuntas, entre muchas otras. Por su parte, los juicios de expertos pueden seguir

diferentes escalas de acuerdo con la extracción de características observables o el uso de la experiencia acumulada.

Cualquiera sea el caso, existen dos nociones subyacentes que en ocasiones se presentan como opuestas y en otros casos como complementarias; la similitud y la disimilitud. En este sentido, para su comprensión se hace necesario una caracterización exacta de las propiedades matemáticas que permiten su definición. En general, se conciben tales nociones a partir de funciones $d, s : X \times X \to \mathbb{R}$, con $X \subseteq \mathbb{R}$, donde la Tabla 1.1 sintetiza las propiedades verificables según sea el caso.

Funciones de disimilitud		Funciones de similitud	
Propiedad	Definición	Propiedad	Definición
Principio mínimo (MIN)	$d(x,x) \leq d(x,y) \wedge d(y,y) \leq d(x,y)$	Principio máximo (MAX)	$s(x,y) \geq s(x,y) \wedge s(y,y) \geq s(x,y)$
No negatividad (NN)	$d(x,y) \geq 0$	No negatividad (NN)	$s(x,y) \geq 0$
Consistencia débil (wC)	$d(x,x) = c_d(x)$	Consistencia débil (Wc)	$s(x,x) = c_s(x)$
Consistencia fuertes (sC)	$d(x,x) = c_d$	Consistencia fuerte (sC)	$s(x,x) = c_s$
Reflexividad (R)	$d(x,x) = 0$	c_s Normalizado (Nc)	$s(x,x) = 1$
Simetría (S)	$d(x,y) = d(y,x)$	Simetría (S)	$s(x,y) = s(y,x)$
No degeneración (D)	$d(x,y) = d(x,x) \vee d(x,y) = d(y,y)$ entonces $x = y$	No degeneración (D)	$d(x,y) = d(x,x) \vee d(x,y) = d(y,y)$ entonces $x = y$
Desigualdad triangular (T)	$d(x,y) + d(y,z) \geq d(x,z)$	Desigualdad triangular inversa (rT)	$s(x,y) + s(y,z) \leq s(x,z)$
Desigualdad ultramétrica (UM)	$d(x,y) + d(y,z) \geq d(x,z)$	Desigualdad ultrasimétrica inversa (rUS)	$\min s(x,y) + s(y,z) \leq s(x,z)$

Se establecen $d(x,x) = c_d(x)$ y $s(x,x) = c_s(x)$ como funciones $(c_d, c_s : X \to \mathbb{R})$. Fuente [86].

Tabla 1.1: *Propiedades matemáticas complementarias verificables en funciones de similitud o disimilitud.*

De forma inmediata se observa que las funciones de disimilitud generalizan el

concepto de distancia, presentándose así la noción de distancia como una función de disimilitud que cumple ciertas condiciones. Por lo tanto, las propiedades presentadas permiten establecer una clasificación tanto de las funciones de similitud como de las funciones de disimilitud de acuerdo al cumplimiento o no de tales propiedades. Por ejemplo, se distinguen dentro de las funciones de disimilitud, la métrica (denominada también distancia) y la cuasi-métrica, por el cumplimiento o no, de la desigualdad triangular respectivamente. Por lo tanto, con base en la Tabla 1.1, a continuación se presenta un listado de la clasificación desarrollada en [86]:

- **Distancia básica**. Si cumple (MIN). Equivalentemente **Similitud básica**, si cumple (MAX).
- **Distancia primitiva**. Si cumple (MIN) y (NN). Equivalentemente **Similitud primitiva**, si cumple (MAX) y (NN).
- **Distancia debilmente consistente**. Si cumple (MIN), (NN) y (wC). Equivalentemente **Similitud debilmente consistente**, si cumple (MAX), (NN) y (wC).
- **Distancia fuertemente consistente**. Si cumple (MIN), (NN) y (sC). Equivalentemente **Similitud fuertemente consistente**, si cumple (MAX), (NN) y (sC).
- **Distancia general o distancia hueca**. Si cumple (MIN), (NN) y (R). Equivalentemente **Similitud general o similitud hueca**, si cumple (MAX), (NN) y (nC).
- **Predistancia o premétrica** . Si cumple (MIN), (NN), (R) y (S). Equivalentemente **Presimilitud**, si cumple (MAX), (NN), (nC) y (S).
- **Cuasidistancia o cuasimétrica**. Si cumple (MIN), (NN), (R), (S) y (D). Equivalentemente **Cuasisimilitud**, si cumple (MAX), (NN), (nC), (S) y (D).

- **Semimétrica** . Si cumple (MIN), (NN), (R), (S) y (T). Equivalentemente **Semisimilitud**, si cumple (MAX), (NN), (nC), (S) y (rT).

- **Distancia o métrica** . Si cumple (MIN), (NN), (R), (S), (D) y (T). Equivalentemente **Similitud o similitud de Minkowsky**, si cumple (MAX), (NN), (nC), (S), (D) y (rT).

- **Ultramétrica** . Si cumple (MIN), (NN), (R), (S), (D) y (UM). Equivalentemente **Ultrasimilitud**, si cumple (MAX), (NN), (nC), (S), (D) y (rUS).

Con base en lo anterior, a continuación se mencionan algunas definiciones ya clásicas, sobre todo en el ámbito de lo borroso que encajan de forma natural en la clasificación planteada anteriormente. Tales definiciones serán empleadas a partir del siguiente capítulo con el fin de establecer procesos de comparación.

Definición 11. *[62] Sea X un conjunto y w un subconjunto borroso de $X \times X$. Entonces w es denominada una relación de proximidad (también, relación de tolerancia) sobre X si se cumplen las siguientes propiedades, $\forall x, y \in X$:*

1. *Reflexiva:* $w(x, x) = 1$
2. *Simétrica:* $w(x, y) = w(y, x)$

Definición 12. *[83] Sea X un conjunto y v un subconjunto borroso de $X \times X$. Entonces v es denominada una relación de similitud sobre X si se cumplen las siguientes propiedades, $\forall x, y, z \in X$:*

1. *Reflexiva:* $v(x, x) = 1$
2. *Simétrica:* $v(x, y) = v(y, x)$
3. *Min-transitiva:* $v(x, z) \geq \min\{v(x, y), v(y, z)\}$

Considerando las definiciones expuestas y en busca de mantener la homogeneidad en el tratamiento de los términos, en adelante se hará referencia a las *relaciones de similitud* en el sentido de la Definición 12, las cuales corresponden con las denominadas funciones de ultrasimilitud. De igual forma, al hacer referencia a las *relaciones de proximidad*, de acuerdo con la Definición 11, se está haciendo alusión a las funciones de presimilitud. Finalmente, al abordar las *funciones de disimilitud*, particularmente se estará haciendo referencia a las denominadas cuasimétricas.

El uso particular de la medidas planteadas en el párrafo anterior, corresponde con la necesidad del cumplimiento de las propiedades específicas para el desarrollo de la propuesta.

A continuación, se exponen los conceptos y resultados correspondientes a la teoría de los subgrupos borrosos y que juntos con lo expuesto sobre operadores de agregación y medidas de similitud y disimilitud, serán empleados en el marco de la clasificación borrosa.

1.3. Subgrupos borrosos

A partir del trabajo pionero de Rosenfeld [97] se ha dado apertura a una amplia variedad de aplicaciones y desarrollos teóricos sobre la noción de subgrupos borrosos, en particular el trabajo desarrollado en [68] establece la integración de los conceptos de subgrupos borrosos y relaciones de similitud. Esta integración conduce, como ocurre en el caso nítido o crisp, a una estrategia de doble vía para la construcción de subgrupos borrosos a partir de relaciones de similitud o viceversa.

A continuación se exponen los conceptos y definiciones necesarias para el desarrollo de la presente tesis, en el marco de la teoría de subgrupos borrosos.

Definición 13. *[83, 97] El conjunto de todos los subconjuntos borrosos de X es*

llamado el conjunto potencia de X y denotado por $FP(X)$. Sea $G = (G, \cdot)$ un grupo arbitrario y sea $\mu \in FP(G)$. Entonces μ es llamado un subgrupo borroso de G si,

1. $\mu(x \cdot y) \geq \min\{\mu(x), \mu(y)\}$, $\forall x, y \in G$ *y*
2. $\mu(x^{-1}) \geq \mu(x)$, $\forall x \in G$ [1]

Ejemplo 1. *Sea $G = \{e, a, b, ab\}$ el grupo cuarto de Klein. El subconjunto borroso de G definido por $\mu(e) = \mu(ab) = 0{,}7$ y $\mu(a) = \mu(b) = 0{,}2$ es un subgrupo borroso de G.*

Definición 14. *[83] Sea G un grupo y sea μ un subgrupo borrosos de G. Entonces μ es llamado un subgrupo borroso normal de G si es un subconjunto borroso Abeliano de G, es decir, $\mu(x \cdot y) = \mu(y \cdot x)$, $\forall x, y \in G$.*

Uno de los primeros elementos que permite establecer la relación entre subgrupos borrosos y relaciones de similitud viene dado por la Proposición 1.

Proposición 1. *[68] Sea v una relación borrosa sobre Ω. v es una relación de similitud sobre Ω si y solo si cada $v_t = \{(x, y) | v(x, y) \geq t\}$ es una relación de equivalencia sobre Ω para todo $t \in [0, 1]$.*

Considerando las relaciones de equivalencia de la Proposición 1, se establece el Teorema 1.

Teorema 1. *[68] Sea v una relación de similitud sobre un conjunto finito G. Entonces existe una operación binaria $\cdot$ sobre G, tal que $G = (G, \cdot)$ es un grupo y μ un subgrupo borroso de G tal que $v(x, y) = \mu(x \cdot y^{-1})$ o $v(x, y) = \mu(x^{-1} \cdot y)$, $\forall x, y, z \in G$, si y solo si, las clases de equivalencia determinadas por la relación de equivalencia crisp $v_t = \{(x, y) | v(x, y) \geq t\}$ tienen el mismo tamaño para todo $t \in [0, 1]$.*

En el contexto de la relaciones de similitud, se establece la propiedad de ser *invariante por traslaciones* a través de la Definición 15.

[1] En una estructura de grupo $(G, *)$, x^{-1} denota el inverso del elemento x para la operación $*$.

Definición 15. *[68] Sea G un grupo y sea v una relación de similitud sobre G. Entonces se dice que v es invariante por derecha si $(\forall x, y, z \in G)\ v(x, y) = v(x \cdot z, y \cdot z)$. De forma similar, se dice que v es invariante por izquierda si $(\forall x, y, z \in G)\ v(x, y) = v(z \cdot x, z \cdot y)$. Se dice que v es invariante por traslaciones si es invariante tanto por izquierda como por derecha.*

Finalmente, de acuerdo con la Definición 15, el Teorema 2 establece la conexión entre una *relación de similitud invariante por traslaciones* y un *subgrupo borroso normal.*

Teorema 2. *[83] Sea v una relación de similitud invariante por traslaciones sobre G. Entonces v es un subgrupo borrosos de $G \times G$ si y solo si μ es un subgrupo borroso de G tal que $\mu(x) = v(e, x)$ y μ es conmutativo, donde e es el elemento identidad de G.*

Los conceptos y resultados de esta sección, se fusionarán y aplicarán a los sistemas de clasificación borrosa con el fin de consolidar una propuesta para la evaluación de particiones borrosas. Por lo tanto, a continuación se presenta el marco de referencia general sobre clasificación borrosa.

1.4. Clasificación borrosa

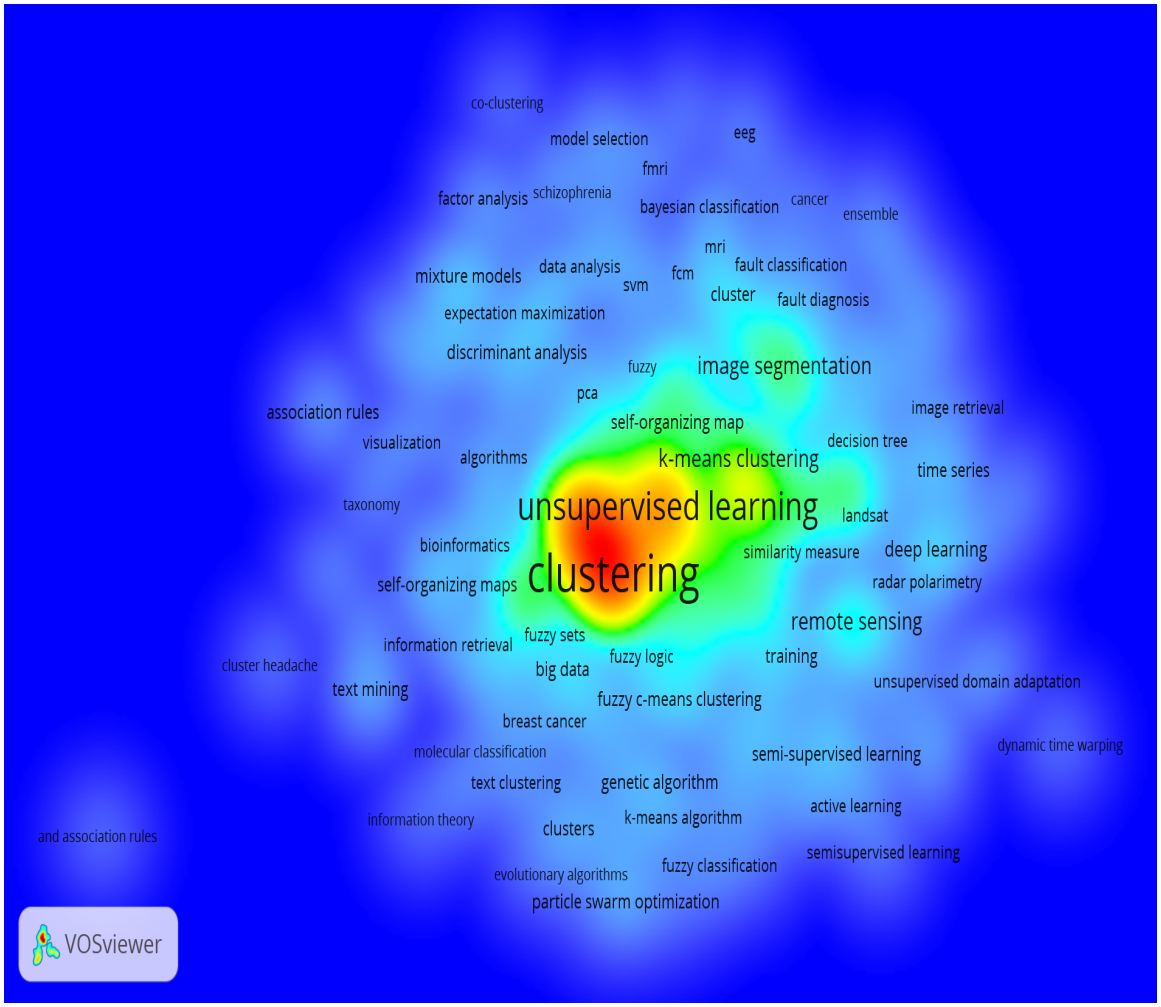
El aprendizaje automático o machine learning como área de investigación tiene por objetivo desarrollar métodos computacionales que sean capaces de "aprender" de acuerdo con la experiencia acumulada. En general, se busca establecer modelos o sistemas aptos que permitan organizar el conocimiento existente o diseñar técnicas que permitan imitar el comportamiento humano o de un experto de manera automática [104].

Una de las ramas con mayores desarrollos dentro del aprendizaje automático ha sido *el reconocimiento de patrones*, la cual ha encontrado *en el análisis clasificatorio borroso o difuso* (clasificación borrosa) un enfoque con valiosas herramientas y aplicaciones durante los últimos cincuenta años. El potencial de dicho enfoque radica en la posibilidad de representar características de la información expresada de forma lingüística, proporcionar una estimación o representación de información ambigua o vaga a través de grados de pertenencia y, dentro de su aplicación particular en tratamiento de imágenes, extraer regiones y características de imágenes poco nítidas o mal definidas [81]. En general, se han convertido en una poderosa herramienta para modelar e imitar procesos cognitivos del ser humano, especialmente aquellos relacionados con aspectos del reconocimiento de patrones [92].

Los subconjuntos borrosos en general, y el proceso de la clasificación borrosa en particular, han tenido un lugar relevante en el desarrollo e implementación de las metodologías y teorías frente al aprendizaje no supervisado. Evidencia de su participación es el amplio desarrollo de producción académica (artículos científicos) generados alrededor del *clustering* como técnica de clasificación no supervisada y la intensidad con la cual la lógica borrosa en sus diversas facetas ha sido protagonista en los últimos 30 años. La Figura 1.1 presenta un mapa de conocimiento[2] que revela la relación entre los conceptos. Como eje central del mapa se establecen los términos de *unsupervised learning* y *clustering*, a partir de ellos se ubican todos los términos con los cuales se presentan coincidencias dentro de la producción académica. Este proceso se establece mediante el rastreo de palabras claves y títulos de los documentos académicos. Conforme más cerca se encuentra un términos a la zona central (zona de unsupervi-

[2]La búsqueda para la generación del mapa de calor (densidad) se realizó a inicios del año 2020 (16/01/2020) en la base de Wos (Web of sciences), bajo la sintaxis "unsupervised classification (AND) clustering", se ha filtrado por artículos científicos y capítulos de libros, junto con la selección de palabras claves. La búsqueda generó un total de 19.432 documentos. Se observa que las palabras *fuzzy, fuzzy sets, fuzzy logic, fuzzy c-means y clustering*, se encuentran en una zona de calor próxima a *clustering y unsupervised learning*. De igual forma, se evidencia una amplia relación con los términos *image segmentation* y *similarity measure.*

sed learning y clustering), mayor cantidad de producción académica involucra tales términos.



Elaboración propia

Figura 1.1: *Relación de densidad entre los conceptos de Unsupervised learning, clustering y fuzzy. Mapa de calor representado con VOSviewer.*

La clasificación borrosa se puede definir como el cálculo del grado de pertenencia de un objeto a una o varias clases en las cuales ha sido particionado el espacio tratado [89]. En otras palabras, la clasificación borrosa es el proceso por el cual se agrupan elementos en conjuntos borrosos a través de una función de pertenencia que caracteriza la intensidad de verificación de una función proposicional borrosa [59].

Los inicios de la clasificación borrosa se hallan soportados en el trabajo desarro-

llado por Bellman, Kalaba y Zadeh [9], en el cual a partir de la noción de conjunto borroso planteada por Zadeh [118] se proponen las operaciones de abstracción y generalización en el reconocimiento de patrones. La abstracción consiste en la estimación de las funciones de pertenencia $f_{A,}f_B : \Omega \to [0, 1]$ de dos conjuntos disyuntos (difusos o no) A y B en un espacio Ω sobre la base del conocimiento de las muestras de n puntos, $\alpha_1...\alpha_n$, que se sabe pertenecen a A y m puntos, $\beta_1...\beta_m$, que se sabe pertenecen a B. Una vez realizada la estimación o diseño de la función f, esta se emplea para calcular los valores de nuevos puntos. En esto consiste la operación de generalización. En particular, este proceso describe lo que se conoce como una tarea de aprendizaje inductivo supervisado de clasificación borrosa.

De acuerdo con Kaufmann. *et. al.* [60], en términos generales la clasificación borrosa puede ser deductiva o inductiva. Es decir, se denomina deductiva cuando se clasifica mediante el uso de funciones de pertenencia predefinidas por seres humanos expertos, como ocurre en el caso de la clasificación de tuplas en una base de datos relacional (ver [79]). En contraste, se denomina inductiva cuando la clasificación se establece a partir del conjunto de datos analizado. En este último caso, la clasificación borrosa puede generarse de forma supervisada o no supervisa. De forma muy general, se puede decir que se considera un enfoque supervisado cuando existe una partición previa del espacio y por ende los datos se encuentran etiquetados en alguna clase. En contraste, bajo un enfoque no supervisado, se parte de un conjunto de datos sin etiquetar en tanto no existe un conjunto de clases preestablecidas. Este último enfoque corresponde con el eje central de la propuesta planteada en la tesis, por lo tanto, de acuerdo con la clasificación dada por [60](ver Tabla 1.2), se profundiza en los modelos referidos a *enfoques no supervisados de clasificación borrosa, conglomerados borrosos*, o *fuzzy clustering*.

Clasificación Borrosa	1	2	3	4	5	6	7	8
Conglomerados borrosos [7, 8]	I	M	A	F	D	L	B,T	Mu
Basados en reglas [2, 33, 49, 53–55, 76, 89]	D,I	M	S	F	Lo	N	B	Mu
Sistemas Neuro-borrosos [80]	I	M	S	G	D	N	B	Mu
Arboles de decisión borrosos [52, 91]	I	M	S	H	Lo	N	T	S
Arboles de clasificación borrosos [32]	I	M	S	H	Lo	N	T	Mu
Arboles con patrones borrosos [103]	I	M	S	H	D	N	B	S
Maquinas de soporte vectorial borroso [31, 71, 72]	I	M	S	F	D	L	T	S
Clasificadores de Kernel borroso [1]	I	M	S	F	D	N	T	Mu
Generación de funciones de pertenencia [3, 63, 85, 90, 110]	D,I	U	S	F	D	N	B	S

Tabla 1.2: *Descripción general de los enfoques existentes para la clasificación borrosa y sus características.*
1. Clasificación: deductivo (D)/inductivo (I) 2. Espacio de entrada: Univariado (U)/Multivariado (M) 3. Modo de entrenamiento: Supervisado (S)/Autónomo (No supervisado) (A) 4. Estructura del modelo: Gráfico (G)/Jerarquía (H)/Función (F) 5. Procesamiento: Lógico (Lo)/Controlado por datos (D) 6. Discriminación de la clase objetivo: lineal (L)/posiblemente no lineal (N) 7. Dirección del modelo: Bottomup (B)/Topdown (T) 8. Salida del modelo: Clase única (S)/Multiclase (Mu). Fuente [60].

1.4.1. Conglomerados borrosos. Fuzzy clustering

Frente al problema de reconocimiento de patrones en un conjunto de objetos desde un enfoque no supervisado, las técnicas de *análisis de conglomerados* o *clustering* se presentan como una alternativa para su tratamiento. Tales técnicas tienen por objetivo la división del conjunto de elementos u objetos en varios grupos o clústeres a través de una medida de similitud. En particular, se busca agrupar los elementos de tal forma que los miembros que pertenecen a un grupo sean lo mas similares posibles, mientras los miembros de diferentes grupos sean lo más disimilar posible.

En principio, las técnicas en el marco de lo nítido o crisp suponen que existen grupos disyuntos en el conjunto de objetos, por lo tanto, la asignación o pertenencia de cada elemento a cada grupo o clúster es única. Sin embargo, en muchos casos los clústeres no cumplen con dicho supuesto y la separación entre los grupos se presenta como una noción borrosa, es decir, cuyos límites no se encuentran claramente defini-

dos. Bajo esta perspectiva, las técnicas de clustering borroso surgen como mecanismos que permiten el tratamiento de esta situación particular, estableciendo la pertenencia de cada objeto o elemento a diversos clústeres a través de grados de pertenencia.

Para obtener la agrupación de los elementos, las técnicas de clustering (nítidas o borrosas), por lo general, se basan en la aplicación de dos tipos de algoritmos; los algoritmos basados en una función objetivo también conocidos como iterativos, o mediante algoritmos jerárquicos. A continuación se exponen algunas generalidades sobre ambos algoritmos para posteriormente, en la Sección 1.4.1.1, profundizar en los algoritmos iterativos borrosos debido al interés particular de la presente tesis; la evaluación de particiones borrosas.

En general, los algoritmos jerárquicos producen un dendrograma que representa la agrupación anidada de patrones y niveles de similitud en los que cambian las agrupaciones. De estos, los algoritmos de enlace único y enlace completo son los más populares. Estos dos algoritmos difieren en la forma en que caracterizan la similitud entre un par de grupos. En el método de enlace único, la distancia entre dos grupos es el mínimo de las distancias entre todos los pares de patrones extraídos de los dos grupos (un patrón del primer grupo y el otro del segundo). En el algoritmo de enlace completo, la distancia entre dos grupos es el máximo de todas las distancias por pares entre patrones en los dos grupos. En cualquier caso, dos grupos se fusionan para formar un grupo más grande basado en criterios de distancia mínima. El algoritmo de enlace completo produce grupos estrechamente unidos o compactos [56].

Por su parte, los algoritmos de agrupación iterativos obtienen una única partición de los datos en lugar de una estructura de agrupación, como el dendrograma producido por una técnica jerárquica. Uno de los problemas más notorios y de mayor interés en el trabajo con este tipo de algoritmos es la *elección del número de grupos de salida deseados*. Las técnicas de partición generalmente producen grupos al optimizar una función de criterio definida localmente (en un subconjunto de los patrones) o

globalmente (definida en todos los patrones). En la práctica, el algoritmo generalmente se ejecuta varias veces, es decir de forma iterativa, con diferentes estados de inicio, y la mejor configuración obtenida de todas las ejecuciones se utiliza como la agrupación de salida.

A continuación se presentan los elementos más relevantes de los algoritmos iterativos borrosos y se expone unos de lo más ampliamente usado, el algoritmo fuzzy c-means.

1.4.1.1. Algoritmos particionales borrosos

Los primeros desarrollos sobre algoritmos particionales borrosos se dan gracias al planteamiento de Zadeh [119] frente a las denominadas *relaciones de similitud*. A partir de tal concepto, se establecen los agrupamientos borrosos basados en relaciones.

En un primer trabajo, Tamura *et al.* [109] establecen un método de clasificación de patrones empleando relaciones borrosas a partir de información subjetiva. Los autores proponen un procedimiento en n pasos a partir de la composición de relaciones borrosas. Se inicia con una relación borrosa reflexiva y simétrica R en sobre un conjunto de elementos X y se muestra que existe un n tal que

$I \leq R \leq R^2 \leq ... \leq R^n = R^{n+1} = ... = R^\infty$ cuando X es un conjunto finito de elementos.

Siendo representadas las relaciones por matrices, I es la matriz $[r_{ij}]$, tal que

$$r_{ij} = \left\{ \begin{array}{ll} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{array} \right\}$$

Con base en lo anterior, R^n es usada para definir una relación de equivalencia R_λ dada por la regla, $R_\lambda(x, y) = 1$ si y solo si, $\lambda \leq R^n(x, y)$; de hecho R^n es una relación de similitud. De igual forma, los datos quedan particionados por R^n.

Posteriormente, Dunn [43] realiza algunas mejoras al algoritmo tanto en tiempo de cálculo como en requisitos de almacenamiento proponiendo un método para calcular R^n.

Específicamente, los algoritmos basados en la función objetivo (o algoritmos de agrupamiento iterativos) determinan para el espacio de características una partición que se busca sea óptima minimizando una función objetivo. Es decir, dado un conjunto de objetos $X = \{x_1, x_2, ..., x_m\}$ no etiquetados y que pertenecen al espacio euclídeo $\mathbb{R}^d$, se propone realizar una partición de estos elementos en un conjunto C de n grupos o conjuntos borrosos con respecto a un criterio dado, es decir, se busca asignar un grado de pertenencia μ_{ik}, con $k = 1, 2, ..., m$, y con $i = 1, 2, ..., n$, a cada elemento de X.

Usualmente, el resultado de la agrupación es representada a través de una matriz de partición U, la cual indica la estructura detectada en el estudio de los datos. Dicha matriz consta de n columnas y m filas, $U = [u_{ik}]$, en la cual, las filas corresponden a la agrupación obtenida. El elemento (i, k) de U indica el grado de pertenencia del objeto k en el grupo o clúster i [92]. Bajo esta perspectiva, surgen de manera natural ciertas restricciones al problema abordado en cuanto a la partición deseada del conjunto de datos. El primero en abordar tal aspecto fue Enrique Ruspini [100] quien establece que para el caso de una partición borrosa se debe cumplir:

1. El total de la pertenencia de los elementos $x_k \in X$ a todas las clases debe ser igual a 1, es decir,

$$\sum_{i=1}^{n} \mu_{ik} = 1, \text{ para todo } 1 \leq k \leq m$$

   por lo que cada objeto x_k puede pertenecer a varias clases, con cierto grado, y el grado total de pertenencia es distribuido entre todas las clases.

2. Cada grupo o clúster construido es no vacía y diferente del conjunto total de

datos, es decir,

$$m > \sum_{k=1}^{m} \mu_{ik} > 0, \text{ para todo } 1 \leq i \leq n$$

En esencia, lo propuesto es una generalización del concepto clásico de partición en conjuntos.

El trabajo propuesto por Ruspini dio paso a la generación de diferentes algoritmos para determinar agrupamientos borrosos. En principio se encontraron algunos fallos con relación al manejo de gran cantidad de datos como lo planteó el mismo Ruspini [101]. Tal problema se mejoró en gran medida tanto de forma teórica como computacional según se expone en [102]. Sin embargo, otra ideas para mejorar el problema relacionado con el volumen de datos lo trabajaron Gitman y Levine en [47], quienes plantean un algoritmo basado en el concepto de conjunto borroso unimodal como equivalente al concepto de α-cortes planteado por Zadeh, el cual establece una partición del conjunto de datos a partir del máximo grado de pertenencia.

En la misma línea, Dunn [41–43], plantea en analogía al algoritmo ISODATA (*Iterative Self-Organizing Data Analysis Techniques*[6]) un algoritmo basado en las particiones borrosas para detectar la presencia o no de agrupamientos compactos y bien separados. Bezdek [12, 13], aplica y avanza en el estudio del algoritmo y posteriormente generaliza lo hecho por Dunn en [14].

Los trabajos desarrollados por Bezdek, cuya esencia se encuentra en la extensión al campo de lo borroso del algoritmo k-medias [74], desembocan quizás en uno de los algoritmos más ampliamente usado y popularizados [56], y cuya aplicación se ha extendido por más de tres décadas; el algoritmo fuzzy c-means.

El desarrollo del fuzzy c-means se encuentra basado en minimizar la función objetivo

$$J_b(X;U,V) = \sum_{i=1}^{n}\sum_{k=1}^{m}\mu_{ik}^{b}||x_k - v_i||_A^2 \tag{1.1}$$

sujeto a la restricción $\sum_{i=1}^{n}\mu_{ik} = 1$ para todo $k = 1, ..., m$.

Para las Ecuación 1.1 se tienen los siguientes parámetros: X corresponde al conjunto de elementos de cardinal m, x_k es el k-ésimo elemento de X, C corresponde al número de clústeres, $\mu_{ik} \in U$ es el grado de pertenencia de x_k al clúster i, $v_i \in V$ es el centro del i-ésimo clúster o vector prototipo del conjunto borroso μ_i con $1 \leq i \leq n$ y, $b \in [1, \infty)$ es un valor que determina el grado de borrosidad de los clústeres resultantes y se denomina parámetro de borrosidad, o exponente de ponderación. En general, cuanto más grande sea b mayor grado de borrosidad se establece entre los clúster, es decir, existe mayor superposición de los clústeres o en otras palabras, sus fronteras son menos nítidas. En el caso contrario, a medida que b se acerca a 1, el agrupamiento tiende a establecer clústeres claramente diferenciados. De acuerdo con lo anterior, el exponente de ponderación controla los grados de pertenencia que comparten los elementos de X entre los clústeres.

Por otro lado, la distancia $||x_k - v_i||_A^2$ está definida a partir de la matriz A y se establece de acuerdo con la Ecuación 1.2,

$$||x_k - v_i||_A^2 = D_{ikA}^2 = (x_k - v_i)^T A(x_k - v_i). \tag{1.2}$$

La partición generada por el algoritmo fuzzy c-means se lleva a cabo mediante la optimización iterativa de la Ecuación 1.1, con la actualización de los grados de pertenencia μ_{ik} y los centros de los clústeres v_i, dados por la Ecuación 1.3 y la Ecuación 1.4 respectivamente,

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{n}\left(\frac{D_{ikA}}{D_{jkA}}\right)^{\frac{2}{b-1}}} \tag{1.3}$$

con $1 \leq i \leq n$ y $1 \leq k \leq m$.

$$v_i = \frac{\sum_{k=1}^{m} \mu_{ik}^b x_k}{\sum_{k=1}^{m} \mu_{ik}^b} \tag{1.4}$$

con $1 \leq i \leq n$.

El algoritmo se detiene cuando se cumple que $\max_{i,k} ||\mu_{ik}^{j+1} - \mu_{ik}^j|| < \epsilon$, donde ϵ es un criterio de parada entre 0 y 1, y j corresponde con el paso de iteración. Debido a que $\sum_{i=1}^{n} \mu_{ik} = 1$ para todo $k = 1, ..., m$, se cumple que $m > \sum_{k=1}^{m} \mu_{ik} > 0$, para todo $1 \leq i \leq n$. A partir de la Ecuación 1.4, se establece que los centros v_i corresponden con la media ponderada de los objetos que pertenecen a un clúster, donde los pesos de ponderación son los grados de pertenencia.

Considerando lo expuesto hasta el momento, el algoritmo fuzzy c-means se expresa de la siguiente forma:

Paso 1. Considere un conjunto de objetos X de cardinal m, $x = \{x_1, ..., x_m\}$.

Paso 2. Fije el número de clústeres n, $2 \leq n \leq m$ y, el criterio de parada ϵ.

Paso 3. Fije un exponente de ponderación (borrosidad) b.

Paso 4. Inicialice la matriz $U_{m \times C}$ con valores aleatorios tal que $\mu_{ik} \in [0, 1]$ y $\sum_{i=1}^{n} \mu_{ik} = 1$.

Paso 5. Calcule los centros de los clústeres v_i empleando la Ecuación 1.4.

Paso 6. Actualice la matriz U de acuerdo con la Ecuación 1.3.

Paso 7. Repita desde el paso 5, actualizando los centros v_i, si $\max_{i,k} ||\mu_{ik}^{j+1} - \mu_{ik}^j||$ es mayor que el criterio de terminación ϵ.

En particular, cuando en el algoritmo se considera la distancia Euclidiana, se emplea la matriz $A = I$, la matriz identidad y, el algoritmo se enfoca en detectar grupos esféricos. De igual forma, un aspecto relevante en el uso del fuzzy c-means es que, aunque en el proceso de optimización iterativo, no existan garantías de un valor

óptimo, las amplias y variadas aplicaciones en las cuales se ha usado el algoritmo demuestran que el método funciona muy bien [11].

Otro de los aspectos que resalta la relevancia del algoritmo, han sido las múltiples modificaciones y variantes que se han planteado alrededor del mismo. Las dos más conocidas son probablemente la modificación de Gustafson-Kessel (ver [50] , ver también [99]) y la modificación basada en distancias relativas a los primeros componentes principales (ver [14]).

En el primero, la distancia Euclidiana se reemplaza por una distancia de Mahalanobis con una matriz de dispersión borrosa que depende de los centros de los clústeres estimados. Para evitar la solución trivial de $u = \frac{1}{n}$ para todo i y j, las restricciones en los volúmenes de los grupos deben definirse antes de la optimización. La modificación de Gustafson-Kessel del FCM (GK-FCM) es adecuada para situaciones donde los clústeres presentan forma elipsoidal [44].

Frente a los desarrollos sobre algoritmos y técnicas de agrupamiento borroso basados en el concepto de partición planteado por Ruspini, en [92] y particularmente en [115] se encuentra una amplia revisión de los avances tanto teóricos como prácticos.

De forma clara, unos de los conceptos fundamentales en el desarrollo de los algoritmos de agrupamiento y en general, en los enfoques de clasificación borrosos, es el concepto de partición del espacio de características. Como se ha visto, la idea plateada por Ruspini ha generado grandes avances y ha sido uno de los referentes fundamentales de la teoría desarrollada, sin embargo, Krishnapuram [66] plantea un tipo de partición mas general que la partición borrosa, denominada partición posibilística y bajo la cuál la partición de Ruspini aparece como un caso particular. Dicha partición, flexibiliza una de las restricciones planteadas en [100] definiendo la partición posibilística como aquella que cumple para cada función de pertenencia $\mu_{ik} \in [0, 1]$ $k = 1, 2, ..., m$, $i = 1, 2, ..., n$ las siguientes condiciones:

1. $\exists i, \mu_{ik} > 0, \forall k$, es decir, cada objeto x_i tiene un grado de pertenencia a al menos

un clúster y los grados de pertenencia de cada elemento a todos clúster no debe cumplir la restricción $\sum_{i=1}^{n}\mu_{ik} = 1$.

2. Cada grupo o clase construida es no vacía y diferente del conjunto total de objetos, es decir,

$$m > \sum_{k=1}^{m}\mu_{ik} > 0, \text{ para todo } 1 \leq i \leq n$$

Bajo este concepto de partición, no todos los patrones en un mismo grupo son equivalentes, algunos patrones son más representativos o típicos que otros. Un patrón es típico si es similar a los otros miembros del grupo al que pertenece y está muy cerca del prototipo (centro del grupo o clase). Un patrón es atípico si está muy alejado del prototipo, por lo tanto es poco representativo del grupo, pero tiene alguna característica que lo hace pertenecer al grupo [67]

En general, las diferencias entre las particiones del espacio de características planteadas por Ruspini y Krishnapuram, junto con una partición convencional, se pueden ver por un lado, considerando la interpretación de la función μ_{ik} y por otro, a través de la flexibilización de la primera condición, es decir, la pertenencia de los objetos a todas las clases de la partición no deben sumar uno [78].

Con la anterior idea en mente, se han propuesto distintas variantes al algoritmo fuzzy c-means. Por un lado, se han desarrollado variantes específicas considerando la flexibilización de Krishnapuram, dentro de las cuales se encuentra, el algoritmo Possibilistic Fuzzy C-Means (PFCM) [64, 65], el algoritmo Possibilistic C-Means (PCM) [66, 67], y el algoritmo fuzzy Possibilistic C-Means (FPCM) [88], los cuales funcionan bastante bien para grandes conjuntos de datos, aunque presenten debilidades en datos con ruido o son sensibles a los valores de inicialización. Por otro lado, se encuentran algunas generalizaciones que incorporan nuevos componentes al algoritmo (ver [116], [117]) y finalmente, se propone un algoritmo fuzzy c-means *condicional* para permitir

la agrupación de vectores de datos en condiciones basadas en los términos lingüísticos. En este caso, cada vector de datos x_k tiene un dato o variable auxiliar asociada y_k. Los vectores de datos se agrupan en condiciones basadas en algunos términos lingüísticos definidos en la variable auxiliar. Estos términos lingüísticos se tratan como conjuntos borrosos (o relaciones cuando se trata de un vector). Por ejemplo, si el vector x_k corresponde a los registros de una base de datos médica, entonces una variable auxiliar puede corresponde a la *edad* y la función de pertenencia tendrá asociado los grados correspondientes a una categoría como *joven*. . Esta última variación es capaz de manejar datos con valores atípicos [70]

La consideración de flexibilizar la partición propuesta por Ruspini, ha sido un aspecto a considerar en la propuesta desarrollada por A. Amo *et al.* [5], referencia fundamental en el desarrollo de la presente tesis. En tal propuesta, se formula un *sistema de partición borrosa* bajo el cual la partición de Ruspini aparece nuevamente como un caso particular. Por su interés particular, dicho sistema se profundiza en la Sección 1.5.

1.4.1.2. Validación de clústeres. Índices

Uno de los mayores desafíos dentro de la clasificación borrosa no supervisada corresponden a establecer mecanismos, a veces deseable que sea único, para validar los resultados obtenidos.

En la búsqueda de algoritmos que puedan identificar de la mejor manera la estructura presente en los datos, la clasificación borrosa no supervisada ha surgido como una alternativa en los problemas en los que los grupos no son completamente disjuntos; específicamente, en la práctica, en muchos casos la separación de grupos es una noción borrosa.

En cualquier caso, surgen varias preguntas interesantes en la implementación de algoritmos de clasificación: ¿Cómo determinar, a partir de un conjunto de algoritmos

dados, aquel que da mejores resultados, con mayor validez, más útiles? O, ¿cómo determinar el número apropiado de clústeres en un conjunto de datos? De hecho, estas preguntas se refieren a cómo es posible validar que la partición resultante es una "buena partición".

En particular, para responder a la segunda pregunta en el caso borroso, se han establecido en la literatura varios índices que miden la borrosidad de los grupos o conglomerados, las estructuras borrosa de los datos, la compacidad intragrupo, la separación, entre otros. Dentro de los índices más utilizados se destacan, por ejemplo, el coeficiente de partición de Bezdek (PC) [12], entropía de partición (PE) [14], coeficiente de partición modificado [35], el índice de Xie-Beni [114], average silhouette width criterion [98], Silueta borrosa (Fuzzy Silhouette) [22].

Aunque los índices de validación son independientes del algoritmo utilizado, varias propuestas han utilizado el algoritmo fuzzy c-means en su aplicación [51, 108, 113]. Sin embargo, cada índice permite evaluar el rendimiento del procedimiento de agrupación de acuerdo con diferentes propiedades, lo que hace muy difícil identificar una sola medida de rendimiento que resuma adecuadamente toda la información relevante para evaluar el resultado de la agrupación.

En general, se considera que un proceso de validación de agrupamientos adecuado requiere la calibración de varias propiedades o características extraídas de los clúster o de la partición obtenida. Por lo tanto, es pertinente establecer un criterio que permita que la partición sea evaluada tanto por criterios *externos* como *internos*. Los criterios internos consideran medidas que evalúan por separado diferentes propiedades o características inherentes a la partición que se está considerando. Por ejemplo, los índices mencionados anteriormente corresponden a criterios internos. En este caso, los índices propuestos permiten evaluar propiedades específicas de la estructura obtenida, y no existe una manera directa de agregarlos a todos en un índice general. Bajo esta perspectiva, corresponde al tomador de decisiones elegir la propiedad o el conjunto

de propiedades que se deben verificar mediante el procedimiento de agrupación.

Los criterios externos generalmente consideran evaluar la estructura de agrupamiento resultante comparándola con una partición independiente de los datos de acuerdo con cierta intuición sobre la estructura de agrupamiento del conjunto de datos [36], o en el caso borroso, comparándola con un partición nítida o crisp. De esta manera, la partición nítida representa el caso ideal en el que, por ejemplo, los bordes de una imagen, es decir, los bordes de los grupos, están claramente definidos.

De acuerdo con [111] la mayoría de índices existentes pueden ser clasificados como aquellos que usan exclusivamente la matriz de los valores de pertenencia y los índices que emplean tanto el conjunto de objetos clasificados como las matriz de pertenencia. A continuación se enuncian algunos de tales índices. Se recuerda la notación $X = \{x_1, ..., x_m\}$ corresponde al conjunto de objetos a clasificar, $U = [u_{ik}]_{m\times n}$ representa la matriz de partición borrosa compuesta por los grados de pertenencia del objeto x_k al clúster i, y $V = \{v_1, ..., v_n\}$ corresponde con los centros de los clústeres.

1. El coeficiente de partición (CP), propuesto por Bezdek [12], indica el promedio relativo correspondiente a la pertenencia compartida entre pares de subconjuntos borrosos en U. Los valores del índice varían en $[\frac{1}{n}, 1]$ y entre más alto se el valor obtenido, mucho mejor. El CP se define como

$$CP = \frac{1}{m}\sum_{i=1}^{n}\sum_{k=1}^{m}\mu_{ik}^2.$$

2. El índice de entropía (IE), propuesto por Bezdek [14], es una medida escalar de la cantidad de borrosidad en U, entre más bajo sea el valor, mucho mejor la partición obtenida. El IE esta dado por

$$IE = -\frac{1}{m}\sum_{i=1}^{n}\sum_{k=1}^{m}\mu_{ik}\log_a \mu_{ik}.$$

3. El coeficiente de partición modificado (CPM), propuesto por [35], busca reducir la tendencia evolutiva monótona que presentan el CP y IE con respecto al número de clústeres n. Los valores del índice oscilan en $[0, 1]$ y al igual que el PC, cuanto más alto sea el valor obtenido, mucho mejor. El CPM está dado aplicando una transformación lineal de PC dada por

$$CPM = 1 - \frac{n}{n-1}(1 - PC).$$

Lo índices anteriores basan su medición en el uso exclusivo de la matriz de pertenencia de la partición. Los siguientes índices emplean adicionalmente el conjunto de datos clasificados.

1. Índice de Xie-Beni (XB), desarrollado en [114], se encuentra centrado en la medición de dos propiedades, compacidad y separación. En particular el numerador de la siguiente expresión mide la compacidad de la partición borrosa y el denominador la fuerza de separación de los clústeres,

$$XB = \frac{\sum_{i=1}^{n}\sum_{k=1}^{m}\mu_{ik}^{b}||x_k - v_i||^2}{m \min_{i,k}||v_i - v_k||^2}$$

2. Índice de Kwon, (Kwon), desarrollado en [69] busca eliminar la tendencia a disminuir monótonamente cuando el número de conglomerados se acerca al número de puntos de datos del índice XB. Para lograr esto, se introdujo una función de penalización en el numerador del índice de Xie y Beni, por lo tanto, el índice de Kwon se define como

$$Kwon = \frac{\sum_{k=1}^{m}\sum_{i=1}^{n}\mu_{ik}^{2}||x_k - v_i||^2 + \frac{1}{c}\sum_{k=1}^{m}||v_i - v_k||^2}{m \min_{i\neq k}||v_i - v_k||^2}$$

Un amplia variedad de índices se han establecido bajo distintas propiedades. Por ejemplo, en [113] se amplía el concepto de compacidad y separación de una partición borrosa. En [46] se aborda el concepto de hípervolumen y densidad y en [95], se profundiza en la variación de dispersión intraclúster. Para los efectos de la presente tesis, se considerarán los índices descritos correspondientes a CP, IE, CPM, XB y Kwon, por tener un uso generalizado y un reconocimiento ya establecido a partir de su abordaje en distintas aplicaciones.

1.5. Sistemas de clasificación borrosa

Los enfoques revisados hasta el momento, aunque han permitido abordar una gran variedad de problemas sobre clasificación y han impulsado de forma notoria los desarrollos tanto teóricos como prácticos, presentan en particular algunas críticas importantes:

1. El concepto de partición borrosa propuesto por Ruspini puede ser entendido en general como un sistema de clasificación borroso deseable. Sin embargo, en diversas aplicaciones reales se convierte en una restricción que los decisores (al menos en una primera etapa) no son capaces de cumplir. De igual forma, es posible que la familia de clases en consideración se encuentre muy lejos de definir dicha partición [4].

2. La partición borrosa tipo rejilla junto con la generación de reglas borrosas para problemas de clasificación de alta dimensionalidad tiene una deficiencia significativa: conforme aumenta la dimensión del espacio, el número de reglas crece de manera exponencial y su desarrollo algorítmico en términos de complejidad y tiempo de ejecución también es demasiado elevado [30], [40].

3. Cai y Kwan [20] establecen que la mayoría de los sistemas existentes basados en reglas borrosas tienen dificultades para derivar reglas de inferencia y funciones de pertenencia directamente de los datos de entrenamiento. Sin embargo, la generación

de funciones de pertenencia a partir de datos tienen varias ventajas como los son: menores costos de configuración para la construcción de sistemas borrosos y la búsqueda y visualización de asociaciones ocultas en los datos utilizando las funciones de pertenencia. Wijayasekara y Manic [112] mencionan que los métodos basados en la generación de una función de pertenencia pueden aumentar la comprensibilidad de los clasificadores borrosos. Sin embargo, como señalan Makrehchi y Kamel [75], generar funciones de pertenencia borrosas a partir de datos reales es una de las cuestiones más difíciles en el diseño de sistemas borrosos.

4. En ocasiones, no se considera que el proceso de clasificación es altamente dependiente de la familia de clases en consideración, lo cual hace que el establecimiento de la función de pertenencia $\mu(x)$ para cada clase se considere como un problema aislado de otras clases.

5. No existe un método óptimo. A mayor complejidad, los problemas requieren del uso simultaneo de enfoques alternativos [4].

A partir de lo anterior, A. del Amo, Montero, Biging y Cutello [5] plantean un sistema borroso de clasificación mediante el empleo de funciones de agregación desde un enfoque recursivo. Es decir, a partir del enfoque clásico de agregación de la información mediante un solo índice (reglas conjuntivas asociativas), se extiende la noción de asociatividad a través del establecimiento de reglas recursivas.

Teniendo como marco de referencia los trabajos desarrollados por Dombi [37], [38] sobre operadores de agregación, se plantea en una primera fase, un enfoque alternativo para aquellos conectivos no asociativos mediante el uso del concepto de *recursividad* [34]. Formalmente se tiene:

Definición 16. *[5] Una regla recursiva ρ es una familia de operadores de agregación*

$$\{\rho_n : [0,1]^n \to [0,1]\}_{n>1}$$

tal que existe una regla de ordenamiento π *y dos secuencias de operadores binarios* $\{L_n : [0,1]^2 \to [0,1]\}_{n>1}$ *y* $\{R_n : [0,1]^2 \to [0,1]\}_{n>1}$ *tal que para cada* n *y para cada* $(a_1, ..., a_n) \in [0,1]^n$, $\rho_n(a_{\pi(1)}, ..., a_{\pi(n)}) = L_n(\rho_{n-1}(a_{\pi(1)}, ..., a_{\pi(n-1)}), a_{\pi(n)}) = R_n(a_{\pi(1)}, \rho_{n-1}(a_{\pi(2)}, ..., a_{\pi(n)})$.

Una regla recursiva es una familia de operadores que permite un ajuste secuencial mediante una aplicación sucesiva de operadores binarios, una vez que los datos se han ordenado correctamente: la regla de ordenamiento asegura que los nuevos datos no introduzcan modificaciones en la posición relativa de los datos ya ordenados. La recursividad asegura la consistencia de dicha familia de operadores al suponer que dicha regla es operativa, en el sentido de que puede evaluarse tanto por izquierda como por derecha mediante una secuencia de operadores binarios.

Una regla recursiva estándar será aquella cuyo orden está dado por la identidad, es decir, la regla de ordenamiento que mantiene el orden de los datos tal como son proporcionados. Por lo tanto, la recursividad de hecho permite la generalización de la asociatividad, es decir, en caso de que dicha secuencia de operadores binarios esté dada por un operador binario único ($L_i = R_j$, para todos i, j), entonces se está haciendo referencia a una regla basada en un operador binario asociativo.

Las reglas recursivas constituyen un mecanismo formal y elegante para tratar agregaciones con número arbitrario de elementos. Además, el enfoque recursivo permite que este mecanismo sea robusto ante cambios en la cardinalidad de los datos, ya que garantiza que todos los operadores en una familia de funciones de agregación estén estrechamente relacionados con un procedimiento de agregación único: los operadores binarios que construyen la regla recursiva (ver también [96]). Lo anterior es útil en el contexto de la clasificación no supervisada, donde el número de clases en las cuales los datos tienen que ser segmentados es típicamente desconocido a priori. En este contexto, las reglas recursivas proporcionan un mecanismo robusto para lidiar con la agregación de información de clases diversas para un número diferente de clases [5],

evaluando automáticamente el rendimiento de la clasificación. A partir de las reglas recursivas se definen entonces los sistemas de clasificación borrosa, en adelante FCS por sus siglas en inglés (fuzzy classifications systems).

Definición 17. *[5] Sea X un conjunto finito de objetos. Un sistema de clasificación borrosa es una familia C de n clases borrosas, donde cada $c \in C$ tiene asociada una función de pertenencia $\mu_c : X \to [0,1]$, junto con una tripleta recursiva (φ, ϕ, N) tal que*

1. *ϕ es una regla recursiva estándar, donde $\phi_2\ (0,1) = \phi_2\ (1,0) = 0$;*
2. *$N : [0,1] \to [0,1]$ es una negación fuerte, es decir, una función continua estrictamente decreciente tal que $N(N(a)) = a$, $\forall\ a \in [0,1]$;*
3. *φ es una regla recursiva estándar tal que, $\forall n > 1$,*
   $\varphi_n(a_1, ..., a_n) = N^{-1}[\phi_n(N(a_1), ..., N(a_n)]$, $\forall (a_1, ..., a_n) \in [0,1]^n$.

Algunos resultados son inmediatos a partir de la consolidación de los FCS. Por un lado, se observa que el operador ϕ es un operador conjuntivo, debido a que $\phi_n(x_1, ..., x_n) = 0$ siempre y cuando exista un $j \in \{1, ..., n\}$ tal que $x_j = 0$. Como consecuencia directa, φ es un operador disyuntivo debido a que $\varphi_n(x_1, ..., x_n) = 1$ siempre y cuando exista un $j \in \{1, ..., n\}$ tal que $x_j = 1$. Por otro lado, como criterio general un FCS establece que un partición borrosa será mejor cuanto más altos sean los valores para φ y más bajos sean los valores para ϕ.

Bajo este marco de referencia, los FCS se propusieron como una estructura que permite la evaluación de la clasificación establecida a partir de las tripletas de De Morgan, es decir, mediante el uso de las reglas recursivas (que satisfacen las leyes de De Morgan) evaluando tres características clave de la familia de clases que se obtienen en una partición borrosa: redundancia, cobertura y relevancia.

La redundancia se refiere a una cierta ortogonalidad en la familia de clases, la cual puede ser vista como un sistema de representación particular del conjunto de

objetos [5]. Es decir, la redundancia sugiere la posible existencia de una representación alternativa que se puede encontrar mediante una redefinición apropiada de la familia de clases o eliminando algunas de las clases existentes. Entonces la redundancia nos habla del grado de coincidencia entre las clases.

La característica de cubrimiento se refiere a cómo una familia de clases verifica completamente los diferentes aspectos de la realidad, o qué tanto han quedado representados los objetos por la partición realizada. Finalmente, la característica de relevancia se entiende como la necesidad de incluir, o no excluir, una clase o familia de clases de la clasificación.

De esta manera, los responsables de la toma de decisiones pueden tener una pista sobre cómo mejorar el rendimiento del clasificador, por ejemplo, buscando algunas clases faltantes, proponiendo clases más grandes o más pequeñas, o eliminando algunas clases.

De acuerdo con lo anterior, los FCS proponen índices que permiten medir el grado de redundancia (superposición) entre clases, el grado en que todas las clases cubren con precisión los aspectos de la realidad en consideración y el grado en que algunas clases pueden ser ignoradas.

Este enfoque es similar a algunos procedimientos estándar en estadísticas y análisis de imágenes, donde se busca una partición nítida de tal manera que cada píxel sea miembro de una y solo una categoría (ver, por ejemplo, [10]). Cualquier proceso de aprendizaje automático también debe tener en cuenta una serie de índices, cada uno de los cuales debe capturar un aspecto específico que debe equilibrarse conjuntamente con fines de aprendizaje.

En esta línea de investigación, algunos primeros enfoques condujeron al desarrollo de trabajos como los presentados en [17–19, 48], y más recientemente en [93, 94] sobre un estudio más profundo de las funciones de agregación para evaluar la redundancia y la cobertura de una clasificación particular. Dichos estudios permitieron el desarrollo

de lo que actualmente se conoce como funciones de solapamiento [18, 48], y funciones de agrupamiento [17, 19]. Estas nuevas funciones no imponen la asociatividad de los operadores más clásicos, como las t-normas y las t-conormas, con las cuales los conceptos de redundancia y cobertura, respectivamente, se definieron originalmente (ver, por ejemplo, [61]).

A continuación se presentan algunos ejemplos de FCS y que serán empleados en los capítulos siguientes. Se han seleccionado de acuerdo a su uso frecuente en algunas aplicaciones.

Ejemplo 2. *Tripleta recursiva asociada a la cardinalidad de conjuntos*

1. $\phi_n(\mu_1(x), \ldots, \mu_n(x)) = \dfrac{3\prod_{k=1}^{n} \mu_k(x)}{1+2\prod_{k=1}^{n} \mu_k(x)}$

2. $\varphi_n(\mu_1(x), \ldots, \mu_n(x)) = \dfrac{1-\left(\prod_{k=1}^{n} (1-\mu_k(x))\right)}{1+2\prod_{k=1}^{n} (1-\mu_k(x))}$

3. $N(\mu(x)) = 1 - \mu(x)$.

La tripleta del Ejemplo 2, corresponde con la generalización del cálculo de la cardinalidad de conjuntos y ha sido empleada en [5], en la aplicación de los FCS.

Ejemplo 3. *Tripleta recursiva del máximo y el mínimo*

1. $\phi_n(\mu_1(x), \ldots, \mu_n(x)) = \min\{\mu_1(x), \ldots, \mu_n(x)\}$

2. $\varphi_n(\mu_1(x), \ldots, \mu_n(x)) = \max\{\mu_1(x), \ldots, \mu_n(x)\}$

3. $N(\mu(x)) = 1 - \mu(x)$.

La tripleta del Ejemplo 3, corresponde a la extensión de las nociones de intersección y unión para subconjuntos borrosos.

Ejemplo 4. *Tripleta recursiva con raíz*

1. $\phi_n(\mu_1(x),\ldots,\mu_n(x)) = \dfrac{\sqrt{\prod_{i=1}^{n}\mu_i(x)}}{\sqrt{\prod_{i=1}^{n}\mu_i(x)+1-\prod_{i=1}^{n}\mu_i(x)}}$
2. $\varphi_n(\mu_1(x),\ldots,\mu_n(x)) = \dfrac{\sum_{i=1}^{n}\mu_i(x)-\prod_{i=1}^{n}\mu_i(x)}{\sqrt{\prod_{i=1}^{n}(1-\mu_i(x))+\sum_{i=1}^{n}\mu_i(x)-\prod_{i=1}^{n}}}$
3. $N(\mu(x)) = 1-\mu(x).$

En la tripleta del Ejemplo 4, el operador ϕ_n corresponde a una función de solapamiento n-dimensional, de acuerdo con lo planteado en [48].

Ejemplo 5. *Tripleta recursiva con negación de Sugeno-Yager*

1. $\phi_n(\mu_1(x),\ldots,\mu_n(x)) = \prod_{i=1}^{n}\mu_i(x)$
2. $\varphi_n(\mu_1(x),\ldots,\mu_n(x)) = \dfrac{\prod_{i=1}^{n}\left(1+2(\mu_i(x))\right)-\prod_{i=1}^{n}(1-\mu_i(x))}{\prod_{i=1}^{n}\left(1+2(\mu_i(x))\right)+\prod_{i=1}^{n}(1-\mu_i(x))}$
3. $N(\mu(x)) = \frac{1-x}{1+2x}$

A través del presente capítulo se han establecido los conceptos y resultados correspondientes con la revisión literaria y, que son necesarios para el desarrollo de la propuesta presentada en la tesis.

A partir del Capítulo 2, se establecen los elementos de construcción propia, producto del proceso investigativo.

Capítulo 2

Similitud (disimilitud) sobre un grupo conmutativo de operadores de agregación

El presente capítulo busca presentar las ideas matemáticas sobre las cuales se cimientan los criterios definidos para la evaluación de particiones borrosas.

A partir de los sistemas de clasificación borrosa (FCS) planteados y desarrollados por [5] (ver también [4]) y que se han revisado en la Sección 1.5, se intentan consolidar ahora las siguientes propuestas: 1) plantear un proceso de agregación sobre los valores obtenidos por los operadores de agregación de un FCS, es decir, definir un valor global de las propiedades evaluadas por cada operador, sobre cada objeto perteneciente a una partición borrosa; 2) considerar una estructura de grupo conmutativo formada por los operadores de agregación y sus negaciones, estudiando algunas propiedades; y 3) aplicar sobre dicha estructura relaciones de similitud y disimilitud que permitan construir procesos de comparación.

Específicamente, el trabajo que se presenta a continuación parte de la hipótesis de que considerar relaciones (de similitud o disimilitud) sobre un grupo conmutativo, permite establecer criterios para determinar la calidad de una partición borrosa y el número óptimo de clases. Dichos criterios, serán aplicables a los grados globales de

las propiedades estudiadas.

2.1. El grado de cubrimiento y solapamiento global

Según se ha descrito en el Capitulo 1, dado un conjunto finito de objetos o elementos X, un FCS se encuentra conformado por un conjunto de clases borrosas C, dos operadores de agregación, uno de tipo conjuntivo ϕ y otro de tipo disyuntivo φ, y una negación estricta N, lo cual conforma la cuaterna (C, ϕ, φ, N). En este sentido, se observa que cualquier FCS queda definido básicamente por el operador ϕ y la negación N, por cuanto $\varphi_n(x_1, ..., x_n) = N^{-1}[\phi_n(N(x_1), ..., N(x_n))]$.

Ahora bien, de acuerdo con cada regla seleccionada, se obtiene un valor agregado para cada $x \in X$ que se interpreta como el grado con el cual la familia de clases borrosas C satisface una propiedad o característica particular para el objeto x (hasta el momento definidos con claridad, los grados de cubrimiento y redundancia). En esta perspectiva, se obtienen tantos valores agregados como objetos existen en X.

Lo anterior plantea una oportunidad por la riqueza de información sobre cada objeto, aunque a su vez genere una necesidad natural; se conoce la información sobre las clases para cada objeto, su grado de cubrimiento o de solapamiento, sin embargo, se desconoce el grado de cubrimiento o solapamiento que las clases de una partición específica están proporcionado sobre todo el conjunto de objetos. Determinar dicho grado permitiría en una primera etapa diferenciar si otra partición es mejor o no para el conjunto de objetos dados.

Por lo tanto, se hace necesario establecer una medida que permita determinar el valor con el cual el conjunto de clases C, satisface una propiedad de interés para todo el conjunto de objetos, es decir, un valor agregado representativo de la propiedad. Por lo tanto, se establece la Definición 18,

Definición 18. *Sea $X = \{x_1, ..., x_m\}$ un conjunto finito de objetos, $C = \{c_1, ..., c_n\}$ una familia de clases borrosas sobre dicho conjunto y ρ_n un operador de agregación*

representando el grado con el cual la familia de n clases borrosas satisface una propiedad específica para un objeto $x \in X$. Entonces, el grado global de ρ_n sobre X, denotado por ρ_n^T, se define como la agregación de los valores de ρ_n para todos los objetos $x \in X$ por medio de un operador de agregación A. Esto es, $\rho_n^T = A(\rho_{1n}, ..., \rho_{mn})$, donde ρ_{jn} denota el grado con el cual las n clases borrosas satisfacen la propiedad específica para cada objeto x_j, $j = 1, \ldots, m$.

Tal operador A puede ser de diferentes clases (conjuntivo, disyuntivo o promedio). La Definición 18, es una generalización de la definición ya presentada en [28] sobre el grado global de cubrimiento. En esta versión se generaliza para cualquier propiedad estudiada por medio de la regla recursiva ρ_n. El Ejemplo 6 ilustra el propósito de la definición dada.

Ejemplo 6. *Sea $C = \{c_1, c_2, c_3\}$ una partición borrosa con tres (3) clases sobre el conjunto $X = \{x_1, x_2, x_3, x_4\}$. Se elije la regla recursiva φ_n del Ejemplo 2 para evaluar el cubrimiento (ver Tabla 2.1)*

Objetos	c_1	c_2	c_3	$\varphi_3(c_1, c_2, c_3)$
x_1	0.8	0.1	0.1	0.6329
x_2	0.7	0.2	0.1	0.5474
x_3	0.3	0.62	0.08	0.5070
x_4	0.5	0.47	0.03	0.4908

Tabla 2.1: *Regla recursiva disyuntiva*

Si se selecciona el operador de agregación tipo promedio

$$A(\varphi_{13}, \varphi_{23}, \varphi_{33}, \varphi_{43}) = \frac{1}{4}(\varphi_{13} + \varphi_{23} + \varphi_{33} + \varphi_{43})$$

(recordar que φ_{ij} denota el grado de cubrimiento de j clases con respecto al objeto i), entonces se obtiene el grado global de cubrimiento del conjunto de clases, es decir, $\varphi_3^T = A(\varphi_{13}, \varphi_{23}, \varphi_{33}, \varphi_{43}) = 0{,}5445$

La Definición 18 se puede hacer extensiva a cualquier propiedad abordada por medio de una regla recursiva. Para lo fines de la propuesta desarrollada, se mostrará la necesidad y pertinencia de emplear operadores tipo promedio cumpliendo la propiedad de ser estables por negaciones fuertes. Una primera aplicación del concepto de *grado global*, se puede encontrar en [25].

2.2. Grupo conmutativo de operadores de agregación para sistemas de clasificación borrosa

A continuación se plantea la estructura sobre la cual se desarrollan la mayoría de resultados del presente documento.

Dado un FCS (C, φ, ϕ, N), se consideran dos nuevas aplicaciones σ_n, $\delta_n : [0,1]^n \to [0,1]$, definidas para todos los operadores de agregación ϕ_n (conjuntivos) y φ_n (disyuntivo). Es decir, se plantea la aplicación $\sigma_n : [0,1]^n \to [0,1]$ definida como,

$$\sigma_n(\mu_1(x), \ldots, \mu_n(x)) = N(\varphi_n(\mu_1(x), \ldots, \mu_n(x))) \tag{2.1}$$

y la aplicación $\delta_n : [0,1]^n \to [0,1]$ definida como

$$\delta_n(\mu_1(x), \ldots, \mu_n(x)) = N(\phi_n(\mu_1(x), \ldots, \mu_n(x))) \tag{2.2}$$

Las aplicaciones definidas adquieren una interpretación natural en el contexto de un FCS. Como se ha descrito en el Capítulo 1, los valores obtenidos a través de los operadores conjuntivos y disyuntivos sobre las clases de una partición borrosa, se interpretan como los grados de solapamiento y cubrimiento, respectivamente. Por lo tanto, al usar la negación fuerte N en la Ecuación (2.1) o en la Ecuación (2.2), los mapeos resultantes puede ser interpretados como los complementos de las proposiciones aplicadas a las clases $\{\mu_1(x), \ldots, \mu_n(x)\}$, es decir, $N(\phi_n(\mu_1(x), \ldots, \mu_n(x)))$ represen-

ta el grado de no redundancia de las clases, entendiendo ϕ_n como una proposición y $N(\phi_n)$ como la negación de tal proposición. De igual forma, $N(\varphi_n(\mu_1(x), \ldots, \mu_n(x)))$ representa el grado de no cubrimiento de las clases.

Desde la perspectiva de la Definición 18, el grado global de no cubrimiento y el grado global de no solapamiento a través de un operador A quedan definidos y se denotan por σ_n^T y δ_n^T.

De igual forma, a partir de la Definición 17 se recuerda que si N es una negación fuerte, entonces

$$\varphi_n(\mu_1(x), \ldots, \mu_n(x)) = N[\phi_n(N(\mu_1(x)), \ldots, N(\mu_n(x)))]$$

y así, dada la aplicación σ_n se cumple que

$$\sigma_n(\mu_1(x), \ldots, \mu_n(x)) = N(N[\phi_n(N(\mu_1(x)), \ldots, N(\mu_n(x)))]$$

y por lo tanto,

$$\sigma_n(\mu_1(x), \ldots, \mu_n(x)) = \phi_n(N(\mu_1(x)), \ldots, N(\mu_n(x))). \tag{2.3}$$

La idea que motiva el planteamiento anterior, se encuentra cimentado en la posibilidad de establecer una relación entre los operadores conjuntivos y disyuntivos, junto con sus negaciones, de tal manera que se pueda evaluar una partición borrosa teniendo en cuenta toda la información posible suministrada por los operadores que conforman un FCS. Esto es, se busca comparar los grados de cubrimiento, solapamiento, no cubrimiento y no solapamiento en un proceso iterativo para particiones borrosas con diferentes números de clases, y determinar la partición con la más alta calidad. Tal proceso iterativo se explicará en el siguiente capítulo.

Es importante resaltar que en el marco de la teoría de operadores de agregación, se tiene una amplia gama tanto de operadores de tipo conjuntivo y disyuntivo como de negaciones fuertes, con lo cual adquiere interés validar el proceso de comparación, independiente de los operadores y la negación fuerte empleada. En este sentido, un hecho que se hace inmediato en el marco de la propuesta aquí planteada, corresponde con la restricción en el uso de operadores de agregación en consonancia con la Definición 8, es decir, la propuesta no admite operadores que cumplan la propiedad de ser *autoduales*. La anterior observación, ya evidente a partir de la caracterización presentada por Silver [105], es relevante reafirmarla por cuanto para operadores autoduales se cumple que $\phi_n = \varphi_n$ para $N = 1 - x$, y con ello, se permitiría que la evaluación de una partición no dependiera de su estructura sino del operador empleado.

Ahora bien, una vez planteadas las dos aplicaciones conformadas por la negación de los operadores, se establece una estrecha relación entre el conjunto de aplicaciones $\phi_n, \varphi_n, \sigma_n$ y δ_n, como se describe a continuación. De acuerdo con la Ecuación 2.1, la Ecuación 2.2 y la Ecuación 2.3, se tiene que los operadores φ_n, σ_n y δ_n pueden ser escritos en términos de ϕ_n, con lo cual, se propone denotar tales operadores como sigue:

- $\phi_n^* = N\Big[\phi_n\Big(N\big(\mu_1(x)\big), \ldots, N\big(\mu_n(x)\big)\Big)\Big] = \varphi_n\big(\mu_1(x), \ldots, \mu_n(x)\big)$,
- $\phi_n^\wedge = \phi_n\Big(N\big(\mu_1(x)\big), \ldots, N\big(\mu_n(x)\big)\Big) = \sigma_n\big(\mu_1(x), \ldots, \mu_n(x)\big)$,
- $\phi_n^\sim = N\Big(\phi_n\big(\mu_1(x), \ldots, \mu_n(x)\big)\Big) = \delta_n\big(\mu_1(x), \ldots, \mu_n(x)\big)$.

Lo anterior implica que "$*$", "$\wedge$" y "$\sim$" son operadores sobre ϕ_n formando φ_n, σ_n y δ_n, es decir, cada uno corresponde con una operación unaria. En este sentido, el operador identidad i puede definirse como $\phi_n^i = \phi_n(\mu_1(x), \ldots, \mu_n(x))$.

Sea B el conjunto de tales operaciones, es decir, $B = \{*, \wedge, \sim, i\}$ y sea $\circ$ la composición de dicha operaciones en el sentido habitual. Esto es, si $\bigtriangledown, \bigtriangleup \in B$, entonces se tiene que $\bigtriangledown \circ \bigtriangleup = \phi_n^{\bigtriangledown\bigtriangleup}$. Por ejemplo, $\phi_n^{\sim\wedge}$ se establece a partir de la composición

de "$\sim$" y "$\wedge$". Como $\phi_n^{\wedge} = \sigma_n$, entonces por la Ecuación 2.3 se cumple que

$$\sigma_n^{\sim} = N(\sigma_n(\mu_1(x), \ldots, \mu_n(x))) = N\left(\phi_n(N(\mu_1(x)), \ldots, N(\mu_n(x)))\right) = \varphi_n.$$

En la misma línea, en términos operacionales se tiene por ejemplo que,

- $\phi_n^{*\sim} = N\Big(\phi_n^*(\mu_1(x), \ldots, \mu_n(x))\Big) = N\Big(N\big[\phi_n(N(\mu_1(x)), \ldots, N(\mu_n(x)))\big]\Big) = \phi_n\Big(N(\mu_1(x)), \ldots, N(\mu_n(x))\Big) = \sigma_n$
- $\phi_n^{*\wedge} = \phi_n^*\Big(N(\mu_1(x)), \ldots, N(\mu_n(x))\Big) = N\Big[\phi_n\big(N(N(\mu_1(x))), \ldots, N(N(\mu_n(x))\big)\Big] = N\Big(\phi_n\big(\mu_1(x), \ldots, \mu_n(x)\big)\Big) = \delta_n$
- $\phi_n^{\wedge\wedge} = \phi_n^{\wedge}\Big(N(\mu_1(x)), \ldots, N(\mu_n(x))\Big) = \phi_n\big(N(N(\mu_1(x))), \ldots, N(N(\mu_n(x))\big) = \phi_n(\mu_1(x), \ldots, \mu_n(x)) = \phi_n$

Sin perdida de generalidad, se hace referencia a ϕ_n para denotar a $\phi_n(\mu_1(x), \ldots, \mu_n(x))$. De igual forma, de ser necesario se emplea la notación $\phi_{kn} = \phi_n(\mu_1(x_k), \ldots, \mu_n(x_k))$, para referirse a la agregación de las funciones de pertenencia μ_n de cada elemento x_k para las n clases. Finalmente, se denota $\phi_{kn}(N(x_k)) = \phi_n(N(\mu_1(x_k)), \ldots, N(\mu_n(x_k)))$ para referirse a la agregación de las negaciones de las funciones de pertenencia μ_n de cada elemento x_k para las n clases.

Considerando el conjunto B y con base en los ejemplos abordados sobre la composición de los operadores unarios, se plantea la Definición 19.

Definición 19. *Sea $G_o = \{\phi_n, \varphi_n, \sigma_n, \delta_n\}$. Una operación binaria $\odot : G_o \times G_o \to G_o$ sobre G_o puede ser definida como, $\theta_n \odot \lambda_n = \phi^{\nabla} \odot \phi^{\triangle} = \phi_n^{\nabla\triangle}$ para todo θ_n, $\lambda_n \in G_o$ y $\nabla, \triangle \in B$.*

Por ejemplo,

$$\delta_n \odot \sigma_n = \phi_n^{\sim} \odot \phi_n^{\wedge} = \phi_n^{\sim\wedge} = \varphi_n.$$

De forma similar, se verifica que

$$\phi_n \odot \varphi_n = \phi_n^{i} \odot \phi_n^{*} = \phi_n^{i*} = \varphi_n$$

Considerando la Definición 19, la Tabla 2.2 presenta los resultados de la operación $\odot$ sobre G_o.

$\odot$	ϕ_n	σ_n	δ_n	φ_n
ϕ_n	ϕ_n	σ_n	δ_n	φ_n
σ_n	σ_n	ϕ_n	φ_n	δ_n
δ_n	δ_n	φ_n	ϕ_n	σ_n
φ_n	φ_n	δ_n	σ_n	ϕ_n

Tabla 2.2: *Operación* $\odot$

Proposición 2. $(G_o, \odot)$ *es un grupo conmutativo.*

Demostración. La demostración es sencilla a partir de la verificación de la Tabla 2.2; ϕ_n es el elemento neutro de la operación y cada elemento es su propio inverso. □

Ahora bien, para el grupo conmutativo conformado $(G_o, \odot)$ es bien sabido que puede ser visto como un subgrupo normal del grupo alternante A_4.[1] Es decir, sea $G_p = \{e, a, b, c\}$ el conjunto de permutaciones de G_o sobre sí mismo, donde

$$e = \begin{pmatrix} \phi_n & \sigma_n & \delta_n & \varphi_n \\ \phi_n & \sigma_n & \delta_n & \varphi_n \end{pmatrix} \quad a = \begin{pmatrix} \phi_n & \sigma_n & \delta_n & \varphi_n \\ \sigma_n & \phi_n & \varphi_n & \delta_n \end{pmatrix}$$

$$b = \begin{pmatrix} \phi_n & \sigma_n & \delta_n & \varphi_n \\ \delta_n & \varphi_n & \phi_n & \sigma_n \end{pmatrix} \quad c = \begin{pmatrix} \phi_n & \sigma_n & \delta_n & \varphi_n \\ \varphi_n & \delta_n & \sigma_n & \phi_n \end{pmatrix}$$

Por lo tanto, bajo la composición de permutaciones, e es la composición identidad y cada composición es su propio inverso. Es decir G_p es un grupo y existe un aplicación entre los grupos $\tau : G_p \to G_o$, donde, $\tau(e) = \phi_n$, $\tau(a) = \sigma_n$, $\tau(b) = \delta_n$ y $\tau(c) = \varphi_n$. Así definida, la aplicación τ es inyectiva y sobreyectiva, es decir, τ es un isomorfismo entre grupos.

[1]El grupo alternante, denotado A_n o $Alt(n)$, es el subgrupo del grupo simétrico S_n en un conjunto de longitud n, es decir, es un grupo de permutaciones. En particular, $A_4 = \{id, (12)(34), (13)(24), (14)(23)\}$.

Hasta este punto, el grupo $(G_o, \odot)$ conformado con, $\phi_n, \varphi_n, \delta_n$ y σ_n se constituye para n clases sobre cada objeto $x \in X$. Es decir, si X tiene m objetos ($|X| = m$), entonces se obtienen m grupos, uno por cada objeto. Por lo tanto, se hace necesario considerar el proceso de agregación que permite trabajar con los grados globales de cada una de las propiedades abordadas a partir de cada operador del FCS.

En este sentido, se requiere que el operador de agregación A (ver Definición 18) seleccionado para tal fin, sea consistente con la estructura de grupo, es decir, se requiere un operador de agregación tal que, al aplicarse sobre los grados de cada propiedad (cubrimiento, solapamiento, no cubrimiento o no solapamiento) para cada clase, los elementos resultantes continúen conformando un grupo. En otras palabras, se requiere que A conserve las operación $\odot$ sobre el conjunto $G_T = \{\phi_m^T, \varphi_m^T, \delta_m^T, \sigma_m^T\}$ y por lo tanto, debe cumplirse que $(G_T, \odot)$ sea un grupo.

De acuerdo con lo anterior, si $|X| = m$ y $|C| = n$, entonces se denotan los correspondientes grados globales de cada una de las propiedades estudiadas por:

- $\phi_m^T = A(\phi_{1n}, \ldots, \phi_{mn})$
- $\varphi_m^T = A(\varphi_{1n}, \ldots, \varphi_{mn})$
- $\sigma_m^T = A(\sigma_{1n}, \ldots, \sigma_{mn})$
- $\delta_m^T = A(\delta_{1n}, \ldots, \delta_{mn})$

Por lo tanto, para ver que $G_T = \{\phi_m^T, \varphi_m^T, \delta_m^T, \sigma_m^T\}$ conserva la estructura de grupo bajo la operación $\odot$ basta con encontrar un operador A que permita escribir cada elemento de G_T en términos de ϕ_m^T conservando la relación establecida entre los elemento del grupo (ver Ecuación 2.1, 2.2 y 2.3).

Se recuerda que la operación $\odot$ se estableció en la Definición 19 y, el propósito a continuación es extender tal operación al conjunto G_T. Es decir, $\odot : G^T \times G^T \to G^T$ puede ser definida como $\theta_m^T \odot \lambda_m^T = (\phi_m^T)^{\nabla} \odot (\phi_m^T)^{\triangle} = (\phi_m^T)^{\nabla\triangle}$ para todo $\theta_m, \lambda_m \in$

G^T y $\bigtriangledown, \bigtriangleup \in B$. Tal extensión es posible siempre y cuando cada $\theta_m^T \in G_T$ pueda ser escrito en términos de ϕ_m^T. Se tiene entonces la Proposición 3.

Proposición 3. *Sea* $G_T = \{\phi_m^T, \varphi_m^T, \delta_m^T, \sigma_m^T\}$, *donde* $\phi_m^T = A(\phi_{1n}, \ldots, \phi_{mn})$, $\varphi_m^T = A(\varphi_{1n}, \ldots, \varphi_{mn})$, $\sigma_m^T = A(\sigma_{1n}, \ldots, \sigma_{mn})$, $\delta_m^T = A(\delta_{1n}, \ldots, \delta_{mn})$. *Si* A *es un operador estable por negaciones fuertes entonces* $(G_T, \odot)$ *es un grupo.*

Demostración. Lo primero es establecer que cada $\theta_m^T \in G_T$ pueda se escrito en términos de ϕ_m^T, para lo cual se tiene:

Para $\varphi_m^T = A(\varphi_{1n}, \ldots, \varphi_{mn}) =$

$$A\Big[N\Big(\phi_{1n}\big(N(\mu_1(x_1)), \ldots, N(\mu_n(x_1))\big)\Big), \ldots, N\Big(\phi_{mn}\big(N(\mu_1(x_m)), \ldots, N(\mu_n(x_m))\big)\Big)\Big].$$

Como A es estable por negaciones fuertes, entonces φ_m^T $= N\Big(A\Big[\phi_{1n}\big(N(\mu_1(x_1)), \ldots, N(\mu_n(x_1))\big), \ldots, \big(\phi_{mn}(N(\mu_1(x_m)), \ldots, N(\mu_n(x_m))\big)\Big]\Big)$.

Se resalta el hecho que el operador ϕ_{kn} se encuentra actuando sobre $N(\mu_1(x_k)), \ldots, N(\mu_n(x_k))$ para cada $x_k \in X$, lo cual previamente se ha denotado como $\phi_{kn}(N(x_k))$, por lo tanto se tiene que $\varphi_m^T = N\Big(A\big(\phi_{1n}(N(x_1)), \ldots, \phi_{mn}(N(x_m))\big)\Big)$ que es equivalente a ϕ_m^T actuando sobre $N(\mu_1(x_k)), \ldots, N(\mu_n(x_k))$. Por lo tanto, φ_m^T se puede escribir en términos de ϕ_m^T. Así pues, se denota $\varphi_m^T = (\phi_m^T)^*$.

Considerando el mismo razonamiento se tiene para σ_m^T que

$\sigma_m^T = A(\sigma_{1n}, \ldots, \sigma_{mn}) = A(N(\varphi_{1n}), \ldots, N(\varphi_{mn})) = N(A(\varphi_{1n}, \ldots, \varphi_{mn}))$, con lo cual, $\sigma_m^T = N\Big(N\Big(A\big(\phi_{1n}(N(x_1)), \ldots, \phi_{mn}(N(x_m))\big)\Big)\Big) = A\big(\phi_{1n}(N(x_1)), \ldots, \phi_{mn}(N(x_m)\big)$. Por lo tanto, σ_m^T ha quedado escrito en términos de ϕ_m^T. Con lo cual, se denota $\sigma_m^T = (\phi_m^T)^\wedge$.

Finalmente, para δ_m^T, se tiene que

$\delta_m^T = A(\delta_{1n}, \ldots, \delta_{mn}) = A(N(\phi_{1n}), \ldots, N(\phi_{mn})) = N(A(\phi_{1n}, \ldots, \phi_{mn})) = N(\phi_m^T)$. Por lo tanto, se denota $\delta_m^T = (\phi_m^T)^\sim$.

Una vez expresados los grados globales, el conjunto $G_T = \{\phi_m^T, \varphi_m^T, \delta_m^T, \sigma_m^T\}$ se puede expresar como $G_T = \{(\phi_m^T)^i, (\phi_m^T)^*, (\phi_m^T)^\wedge, (\phi_m^T)^\sim\}$, donde el operador i puede

definirse como, $(\phi_n^T)^i = A(\phi_{1n}, \ldots, \phi_{mn})$. Por lo tanto, "$*$", "$\wedge$" y "$\sim$" son operadores sobre ϕ_n^T formando φ_n^T, σ_n^T y δ_n^T, es decir, cada uno corresponde con una operación unaria.

Así pues, la operación binaria $\odot$ sobre G_T determina los resultados presentados en la Tabla 2.3

$\odot$	ϕ_m^T	σ_m^T	δ_m^T	φ_m^T
ϕ_m^T	ϕ_m^T	σ_m^T	δ_m^T	φ_m^T
σ_m^T	σ_m^T	ϕ_m^T	φ_m^T	δ_m^T
δ_m^T	δ_m^T	φ_m^T	ϕ_m^T	σ_m^T
φ_m^T	φ_m^T	δ_m^T	σ_m^T	ϕ_m^T

Tabla 2.3: *Operación $\odot$ sobre G_T*

A partir de los resultados de la Tabla 2.3 se verifica de forma directa que ϕ_m^T es el elemento neutro del grupo y cada elemento es su propio inverso. Con lo cual $(G_T, \odot)$ es un grupo.

□

El planteamiento hasta ahora expuesto permitirá consolidar la implementación de relaciones de similitud sobre el grupo conmutativo propuesto, anidando la idea de subgrupos borrosos, lo cual dará sentido y amplitud a la comparación entre elementos del grupo. El resultado presentado en la Proposición 3 permitirá aplicar los resultados de la siguiente sección, tanto al grupo formado por lo operadores de un FCS para cada elemento como al grupo formado por los grados globales.

2.3. Las nociones de similitud y disimilitud sobre el grupo conmutativo

A partir de la estructura de grupo conmutativo establecida con la Definición 19 sobre G_o y con la Proposición 3 sobre G_T, es decir, la estructura conformada por los grados de cubrimiento, solapamiento, no cubrimiento y no solapamiento y los grados

globales correspondientes, se busca establecer un proceso de comparación que permita de forma simultanea, considerar todo los elementos del grupo. Dicha comparación se plantea de tal forma, que más allá de comparar los grados obtenidos, se compare el significado de dichos valores.

Una de las ideas que surgen de manera natural en la búsqueda de un proceso de comparación, es establecer algún tipo de convergencia o divergencia entre los elementos a comparar; ¿qué tan *cercanos* o *distantes* se encuentran los elementos entre sí de acuerdo al marco de referencia o variable que se están contrastando? En este sentido, la consideración de algún tipo de *distancia* que permita comparar los elementos del grupo emerge de forma natural. Sin embargo, la noción básica de distancia aplicada a la estructura de grupo conmutativo, en principio, desconoce dicha estructura. Es decir, en términos de distancia habitual, la noción de grupo establecida no juega un papel relevante o viceversa. Lo anterior, por cuanto se trabaja sobre los valores obtenidos por los operadores como números reales y no como elementos del grupo. Por lo tanto, surge la necesidad de establecer algún mecanismo que permita conjugar ambas nociones. Se considera entonces que las funciones de disimilitud aportarán en tal propósito.

Por otro lado, la intuición sobre ¿qué tan *iguales* son los elementos tratados?, es otro camino que permite establecer un proceso de comparación. De nuevo, la noción básica de igualdad, en términos de la estructura de grupo conmutativo, difícilmente contribuye en la comparación de los elementos, más allá de la relación evidente entre números reales. En este sentido, las relaciones de similitud pueden aportar, bajo ciertas premisas, en la consolidación de un proceso amplio de comparación.

En el desarrollo de la presente sección se abordan tanto la noción de similitud como disimilitud sobre el grupo conmutativo. Para tal desarrollo, necesariamente se acuden a herramientas adicionales que permitan la aplicación de relaciones de similitud o funciones de disimilitud, manteniendo por un lado las propiedades del

grupo establecido o alguna estructura subyacente y, por el otro, la interpretación de los valores correspondientes a cada operador de agregación.

En el contexto borroso, las relaciones de similitud (Definición 12) se definen sobre cualquier conjunto X y dicha relación es vista como un subconjunto borroso de $X \times X$ donde los grados de pertenencia representan el grado similitud entre los elementos de X.

En el caso particular del grupo conmutativo establecido en la sección anterior, cuyos elementos son operadores de agregación y en últimas, valores en el intervalo $[0, 1]$, las relaciones de similitud pueden ser vistas como funciones de $[0, 1]^2$ en $[0, 1]$ y, por lo tanto, se comprenden como un caso particular de una medida de similitud restringida a $[0, 1]$, en consonancia con la clasificación dada en [86] (ver Tabla 1.1), con la denominada *Ultrasimilitud*.

De acuerdo con lo anterior, son pertinentes algunas observaciones: 1) en la siguiente sección se hará referencia al término relaciones de similitud debido a la generalidad de los conceptos abordados y aplicables a cualquier conjunto X. 2) La dualidad entre las medidas de similitud y las medidas de disimilitud queda plasmada de forma natural a través de las definiciones abordadas en [86] y las propiedades plasmadas en la Tabla 1.1, pero más aún, en el marco de funciones de $[0, 1] \times [0, 1]$ en $[0, 1]$ y considerando las denominadas cuasimétricas y cuasisimilitud, entonces se tienen los resultados expresados en las Proposiciones 4 y 5

Se recuerda que las cuasimétricas y las cuasisimilitudes, cumplen las propiedades de: 1) principio mínimo o máximo respectivamente, 2) No negatividad, 3) reflexiva (en el caso de la similitud, la normalización para las funciones definidas en $[0, 1]$ es equivalente a la reflexividad), 4) simétrica y 5) no degeneración.

Proposición 4. *Sea $s : [0, 1] \times [0, 1] \to [0, 1]$ una cuasisimilitud y N una negación fuerte, entonces la función $\bar{d} = N \circ s$ es una cuasimétrica.*

Demostración. En el caso del principio del máximo, la no negatividad y la no degeneración son inmediatas por ser N una negación fuerte, la definición del dominio de las funciones y la reflexividad de s respectivamente. La reflexividad de $\bar{d}$ se evidencia considerando $\bar{d}(x,x) = N(s(x,x)) = N(1) = 0$, para todo $x \in [0,1]$. Para la simetría se tiene que, si $s(x,y) = s(y,x)$ entonces $\bar{d}(x,y) = N(s(x,y)) = N(s(y,x)) = \bar{d}(y,x)$. □

Proposición 5. *$d : [0,1] \times [0,1] \to [0,1]$ una cuasimétrica y N una negación fuerte, entonces la función $\bar{s} = N \circ d$ es una cuasisimilitud*

Demostración. Por la propiedad involutiva de la negación fuerte $N(N(x)) = x$, la demostración es análoga a la demostración de la Proposición 4. □

La Proposición 4 y la Proposición 5 corresponden con casos particulares del trabajo desarrollado en [87], en el cual se extiende el concepto de dualidad sobre ultrasimilitudes y ultramétricas definidas sobre cualquier intervalo de los números reales.

Para el caso del presente trabajo, se considera la dualidad entre cuasisimilitudes y cuasimétricas, por ser estas últimas las empleadas en el desarrollo de la propuesta.

2.3.1. Relaciones de similitud para el grupo de operadores

Establecida la estructura de grupo conmutativo a partir de los operadores de agregación, la idea de relaciones de similitud surge de forma natural si se considera establecer un proceso de comparación, pero adicionalmente la idea se ajusta a la estructura de grupo si se acude a la Definición 13 sobre subgrupos borrosos. En particular, sobre este último hecho, la propiedad de *min*-transitividad de las relaciones de similitud evidencia una estrecha relación (coincidencia) con la definición de subgrupos borrosos.

En esta sección se comprueba que, dado un grupo G definir bajo ciertas condiciones una relación de similitud v sobre $G \times G$, es equivalente a encontrar los subgrupos

borrosos normales de G. Específicamente, si $\mu(x)$ denota la función de pertenencia de un subgrupo borroso de G, entonces se cumple que $\mu(x) = v(e, x)$, donde e es el elemento neutro del grupo. Bajo tal equivalencia queda determinado un proceso de comparación dado por la ordenación de tales subgrupos normales por inclusión. Un resultado clásico de la teoría de grupos.

Por otro lado, a partir de la relación de similitud v, es posible definir relaciones de equivalencia v_t, nuevamente sobre G para todo $t \in [0, 1]$. Tales relaciones de equivalencia pueden ser naturalmente ordenadas por inclusión.

Las relaciones de equivalencia se encuentran compuestas por la información proveniente de la comparación de los operadores del sistema de clasificación borrosa y sus negaciones. Por lo tanto, se puede establecer un orden sobre dicha información y por ende analizar, de manera simultánea, tanto la cobertura como la redundancia de las clases y sus complementos.

Se mostrará que tanto el orden de las relaciones de equivalencia como el orden sobre los subgrupos normales generan un mismo proceso de comparación sobre el grupo G y, por lo tanto, son vías equivalentes para establecer dicho proceso.

El desarrollo de un proceso de comparación en el cual se vinculen las propiedades evaluadas con los operadores del FCS, es un aspecto a destacar en la propuesta desarrollada por cuanto se considera un procedimiento de medición y calibración para diferentes índices, y no solo los índices por separado o incluso un solo índice, como lo sería considerar únicamente el grado de solapamiento o el grado de cubrimiento o su evaluación por separado.

Plantear dos vías equivalentes para establecer el proceso de comparación implica la existencia de dos posibilidades en términos operacionales; por un lado, encontrar los subgrupos borrosos normales o por otro lado, construir la relación de similitud. En este sentido, se presentan las razones y el procedimiento para seleccionar la segunda alternativa.

A partir del marco teórico desarrollado en [97] y [68] (ver también [83]), se consideran los grupos G_o y G_p abordados en la Proposición 2. Sea v_o una relación de similitud sobre G_o y v la relación de similitud sobre G_p definida como sigue:

$$v(x,y) = \text{mín}\left\{v_o\Big(x(\theta), y(\theta)\Big) | \theta \in G_o\right\}, \forall x, y \in G_p \tag{2.4}$$

En términos operacionales se tiene que,

$$v(a,b) = \text{mín}\left\{v_o\Big(a(\phi_n), b(\phi_n)\Big), v_o\Big(a(\sigma_n), b(\sigma_n)\Big), v_o\Big(a(\delta_n), b(\delta_n)\Big), v_o\Big(a(\varphi_n), b(\varphi_n)\Big)\right\}$$

En [83] se prueba que, dado un conjunto Ω y un grupo G de permutaciones de Ω sobre si mismo, la relación de similitud v, construida de acuerdo con la Ecuación 2.4, es una relación de similitud invariante por derecha a través de la relación de similitud v_o.

Para el caso particular de G_o y G_p, se cumple la Proposición 6.

Proposición 6. *Si v_o es una relación de similitud invariante por traslaciones sobre G_o, entonces v es una relación de similitud invariante por translaciones sobre G_p.*

Demostración. Se debe probar que $v(zx, zy) = v(x, y)$ y $v(xz, yz) = v(x, y)$. Entonces se tiene que, $v(zx, zy) = \text{mín}\{v_o(zx(\theta), zy(\theta)) | \theta \in G_o\}$, pero como v_o es invariante por traslaciones, entonces se cumple que $v(zx, zy) = \text{mín}\{v_o(x(\theta), y(\theta)) | \theta \in G_o\} = v(x, y)$. De forma análoga, se cumple que $v(xz, yz) = v(x, y)$ debido a que v_o es invariante por traslaciones. □

En la misma línea, el Teorema 1 y el Teorema 2 ponen de manifiesto los siguientes puntos:

- Cada relación de similitud v sobre un grupo G tiene asociada un subgrupo borroso μ de G, tal que $\mu(x) = v(e, x)$.
- Dado un subgrupo borroso sobre G, existe una relación de similitud v definida por $v(x, y) = \mu(xy^{-1})$.
- Una relación de similitud v definida a partir de un subgrupo borroso es invariante por traslaciones si y solo si el subgrupo borroso es conmutativo.

Con base en lo anterior, se prueba la Proposición 7.

Proposición 7. *Sea G un grupo. Si v es un subgrupo borroso de $G \times G$ y v es una relación de similitud invariante por traslaciones sobre G, entonces v es un subgrupo borroso normal de $G \times G$.*

Demostración. Trivial si G es conmutativo. Por otro lado, por el Teorema 2, existe una subgrupo borroso conmutativo μ de G tal que $\mu(x) = v(e, x)$, por lo tanto, $\mu(xy^{-1}) = v(e, xy^{-1})$ y como v es invariante por traslaciones, entonces $\mu(xy^{-1}) = v(y, x) = v(x, y)$. Ahora, como v es un subgrupo borroso de $G \times G$, v será normal siempre que v sea conmutativo. Por lo tanto, se tiene que $v((x_1, y_1)(x, y)) = v(x_1x, y_1y) = \mu(x_1x(y_1y)^{-1}) = \mu(x_1xy^{-1}\ y_1^{-1})$ y como μ es conmutativo, se cumple que $\mu(x_1xy^{-1}y_1^{-1}) = \mu(xx_1y^{-1}y_1^{-1}) = v(xx_1, yy_1) = v((x, y)(x_1, y_1))$.

□

La Proposición 7 establece que dado un grupo G es posible establecer una relación de similitud invariante por translaciones, de tal forma que v es a su vez un subgrupo borroso normal de $G \times G$. Por lo tanto, se deduce que:

1. Las clases laterales a izquierda y las clases laterales a derecha obtenidas por la relación de similitud sobre G, de acuerdo con el Teorema 1, son las mismas.
2. Los subgrupos borrosos normales de G forman un retículo bajo inclusión.

En general, se puede mostrar que el orden dado por la contenencia de las relaciones de equivalencia definidas a partir de v (ver Teorema 1), es el mismo que el orden dado por la contenencia de los subgrupos borrosos normales. Para ello se presenta la Proposición 8, que plantea un resultado aún más general.

Proposición 8. *Sea $(G, \cdot)$ un grupo y sea μ un subgrupo borroso de G. Para cada $t \in [0,1]$ sea $u_t = \{x \in G : \mu(x) \geq t\}$. Sea a y $b \in [0,1]$ tal que $a \leq b$. Entonces $u_b \subseteq u_a$.*

Demostración. Sea x en u_b, entonces $\mu(x) \in [b,1]$. Si $a \leq b$ entonces, $[b,1] \subseteq [a,1]$. Por lo tanto, $\mu(x) \in [a,1]$ es decir, $\mu(x) \geq a$ y así, $x \in u_a$.

□

En particular, dado que tanto el grupo G_o como G_T son conmutativos, los subgrupos borrosos obtenidos son normales.

Con los resultados presentados hasta ahora se deduce que establecer una relación de similitud sobre grupo un G es equivalente a encontrar subgrupos borrosos normales de G. Por lo tanto, se cuenta con una vía para construir una relación de similitud sobre G. Sin embargo, tal construcción en general desemboca en la falta de interpretación de dicha relación $v(x,y)$, es decir, la relación, aunque consistente, carece de significado práctico o interpretabilidad en el contexto de un FCS.

Lo anterior, se manifiesta de la siguiente forma: si la relación v está construida vía un subgrupo borroso, entonces la relación aplicada sobre dos elementos del grupo no depende del valor que cada elemento representa como resultado del operador de agregación correspondiente. Por ejemplo, sea $G_o = \{\phi_n, \varphi_n, \sigma_n, \delta_n\}$ y sea μ^* un subgrupo borroso de G_o de tal forma que $\mu^*(\phi_n) = 1$, $\mu^*(\varphi_n) = 0,2 = \mu^*(\delta_n)$ y $\mu^*(\sigma_n) = 0,3$. Por otro lado, sea $\mu^{\odot}$ otro subgrupo borroso de G_o de tal forma que $\mu^{\odot}(\phi_n) = 1$, $\mu^{\odot}(\varphi_n) = 0,1 = \mu^{\odot}(\delta_n)$ y $\mu^{\odot}(\sigma_n) = 0,9$. Claramente, μ^* y $\mu^{\odot}$ determinan dos relaciones de similitud $v_*(x,y)$ y $v_{\odot}(x,y)$ que no dependen de los valores de cada elemento de G_o.

De acuerdo con lo anterior, adquiere relevancia indagar por la construcción del proceso de comparación por la vía del planteamiento de la relación de similitud. Al final, la relación obtenida define a su vez subgrupos borrosos normales y relaciones de equivalencia, sin perder la finalidad de un ordenamiento por inclusión. Y adicionalmente, la similitud entre elementos, puede dotar de sentido la interpretación del ordenamiento hallado.

Para establecer el proceso de construcción de la relación v sobre un grupo conmutativo G, se debe tener en cuenta que tal relación, de acuerdo con la Ecuación 2.4, debe cumplir la propiedad de ser invariantes por translación, gracias a que dicha relación v es consistente con la estructura de grupo establecida. En este sentido, se resalta que la construcción de la relación de similitud v requiere de la relación v_0.

Sin embargo, establecer una relación de similitud no siempre es un proceso simple debido a la complejidad en la verificación de la propiedad de *min*-transitividad. Por lo tanto, para obtener una relación de similitud v_o sobre G, la cual permita generar la relación de similitud v vía la Ecuación 2.4, se propone considerar una relación de proximidad denotada por w. Tal relación asegura que en el proceso de comparación, cada elemento de G se declare como totalmente similar a sí mismo, es decir, la similitud del elemento consigo mismo es igual a 1. De igual forma, la condición de simetría de las relaciones de proximidad asegura un proceso de comparación consistente debido a que las diferencias o similitudes entre dos elementos no deberían depender del orden en que se relacionan, o por lo menos si se trabaja con valores agregados en $[0, 1]$. En la misma línea, las propiedades de reflexividad y simetría son muy apropiadas para expresar el grado de *cercanía* o *proximidad* entre los elementos.

Ahora bien, para garantizar la propiedad de *min*-transitividad, se propone calcular la clausura transitiva de w. Precisamente, se recuerda que la propiedad transitiva distingue relaciones de proximidad de relaciones de similitud; la clausura transitiva de relaciones de proximidad determinan relaciones de similitud.

De acuerdo con lo anterior, planteada w, que en términos generales es más sencilla de proponer, se denota por $\widetilde{w}$ la clausura transitiva de w, con lo cual, $\widetilde{w}$ es la relación de similitud buscada. Es decir, $\widetilde{w} = v_o$. El Ejemplo 7 presenta la construcción de relación particular v para el grupo G_T, y la relación de proximidad considerada.

Ejemplo 7. *Dado el grupo conmutativo* $G_T = \{\phi_m^T, \varphi_m^T, \sigma_m^T, \delta_m^T\}$, *se define el operador* $\xi : G_T \times G_T \to [0,1]$ *tal que para todo* θ_m^T, $\lambda_m^T \in G_T$ *se tiene que* $\xi(\theta_m^T, \lambda_m^T) = \left|\theta_m^T - \lambda_m^T\right|$. *Se recuerda que cada* θ_m^T *está definido a partir del operador* A *estable por negaciones fuertes y dado por la Definición 18, es decir* $\theta_m^T = A(\theta_{1n}, \ldots, \theta_{mn})$. *Para el caso particular, sea*

$$A = \frac{1}{m}\sum_{k=1}^{m}\theta_{kn}$$

Como se mencionó anteriormente, $\theta_{kn} = \theta_n(\mu_1(x_k), ..., \mu_n(x_k))$ *consiste en la agregación de las funciones de pertenencia* μ_n *de los elementos* $x_k \in X$ *para las* n *clases. Ahora, considerando* ξ *se selecciona la relación de proximidad* w *dada por*

$$w(\theta_m^T, \lambda_m^T) = 1 - \xi(\theta_m^T, \lambda_m^T) = 1 - \left|\theta_m^T - \lambda_m^T\right|$$

donde $m = |X|$. *Una vez establecida* w, *se calcula su clausura transitiva para obtener,* v_o, *es decir,* $\widetilde{w} = v_o$. *Por lo tanto, de acuerdo con la Ecuación 2.4 se considera la relación de similitud invariante por traslaciones dada por,*

$$v(x,y) = \text{mín}\{v_o(x(\theta_m^T), y(\theta_m^T))|\theta_m^T \in G_T\}$$

donde $x, y \in G_p = \{e, a, b, c\}$.

Es importante resaltar que la selección de la relación de proximidad, debe efectivamente reflejar la idea de cercanía.

En este punto, se resaltan dos aspectos sobre la línea de razonamiento anterior: 1) La relación de similitud v permite emprender un proceso de comparación mediante la

inclusión de relaciones de equivalencia. Con respecto a este proceso, la interpretación de la relación de proximidad es la siguiente: $w(\varphi_m^T, \phi_m^T)$ proporciona el grado que cubre la redundancia de las clases, $w(\varphi_m^T, \sigma_m^T)$ proporciona el grado de proximidad entre el cubrimiento y el no cubrimiento y $w(\varphi_m^T, \delta_m^T)$ proporciona el grado de cubrimiento sin considerar la redundancia de las clases. Por lo tanto, es deseable, por ejemplo, que el grado de similitud entre ϕ_m^T y φ_m^T, es decir, el grado que cubre la redundancia de las clases, sea bajo y menor que el grado de similitud entre φ_m^T y δ_m^T, es decir, el grado de cubrimiento sin considerar la redundancia. Se considera que estas ideas pueden ayudar a establecer un criterio para determinar cuándo una partición es de calidad y cuándo es una partición de alta calidad.

2) La relación de similitud v genera una partición en G_p. Sin embargo, como v se ha construido a partir de w y G_p se ha construido a partir de G_T, en este proceso los valores de $\phi_m^T, \varphi_m^T, \sigma_m^T$ y δ_m^T dados por la relación de proximidad pueden permutar y, por lo tanto, cambiar su significado. Pese a lo anterior, bajo ciertas condiciones es posible garantizar la coherencia de este razonamiento, como se demostrará en el Teorema 3 y el Teorema 4 a continuación.

Para el planteamiento del siguiente teorema, se recuerdan algunos elementos; $G_p = \{e, a, b, c\}$ es el grupo formado por el conjunto de permutaciones pares de G_T sobre si mismo, $v(x, y) = \text{mín}\{v_o(x(\theta_m^T), y(\theta_m^T)) | \theta_m^T \in G_T\}$, $\forall x, y \in G_p$ y v_o es la clausura transitiva de la relación de proximidad w.

Teorema 3. *Considere los grupos* $G_T = \{\phi_m^T, \varphi_m^T, \sigma_m^T, \delta_m^T\}$ *y* $G_p = \{e, a, b, c\}$. *Sea* w *una relación de proximidad sobre* G_T. *Sean,* $1 = w(\varphi_m^T, \varphi_m^T)$, $k = w(\varphi_m^T, \phi_m^T)$, $r = w(\varphi_m^T, \delta_m^T)$, $p = w(\varphi_m^T, \delta_m^T)$ *y* $s = w(\varphi_m^T, \sigma_m^T)$. *Si* $r > s \geq k \geq p$, *entonces* $v(c, b) = r$ *y* $v(c, a) = s$.

Demostración. La Tabla 2.4 presenta los valores de la relación de proximidad w según se han establecido en el teorema.

Sea H la matriz de pertenencia de w de acuerdo con la Tabla 2.4. A partir del algo-

w	ϕ_m^T	σ_m^T	δ_m^T	φ_m^T
ϕ_m^T	1	r	p	k
σ_m^T	r	1	k	s
δ_m^T	p	k	1	r
φ_m^T	k	s	r	1

Tabla 2.4: *Relación de proximidad w sobre G_T*

ritmo expuesto en [62] para el cálculo de la clausura transitiva, entonces se determina $H' = H \cup (H \circ H)$ con el operador máx para la unión y la composición máx $-$ mín para encontrar v_o, es decir, la clausura transitiva de w. Si $H' \neq H$ entonces se elije $H = H'$ y se repite el proceso hasta que se cumpla la igualdad. Aplicado el primer paso, se obtiene

$$H' = \begin{pmatrix} 1 & r & k & s \\ r & 1 & s & s \\ k & s & 1 & r \\ s & s & r & 1 \end{pmatrix}$$

Se consideran entonces los siguientes casos: 1) si $r > s = k = p$, entonces $H' = H$, y por lo tanto, H' es la matriz de pertenencia de v_o. Entonces, se calcula $v(c, b)$ y $v(c, a)$ como sigue:

$$\begin{aligned} v(c,b) = &\text{mín}\left\{v_o\Big(c(\phi_m^T), b(\phi_m^T)\Big), v_o\Big(c(\sigma_m^T), b(\sigma_m^T)\Big), v_o\Big(c(\delta_m^T), b(\delta_m^T)\Big), v_o\Big(c(\varphi_m^T), b(\varphi_m^T)\Big)\right\} = \\ &\text{mín}\left\{v_o(\sigma_m^T, \phi_m^T), v_o(\phi_m^T, \sigma_m^T), v_o(\delta_m^T, \varphi_m^T), v_o(\varphi_m^T, \delta_m^T)\right\} = \\ &\text{mín}\{r, r, r, r\} = r. \end{aligned}$$

De forma análoga se obtiene $v(c, a) = s$. Por otro lado, 2) si $r > s = k > p$ entonces $H' \neq H$, y por lo tanto, el proceso se repite con $H = H'$, es decir, $H'' =$

$H' \cup (H' \circ H')$. Con lo cual se obtiene,

$$H'' = \begin{pmatrix} 1 & r & s & s \\ r & 1 & s & s \\ s & s & 1 & r \\ s & s & r & 1 \end{pmatrix}$$

En este caso, $H'' = H'$ y por lo tanto, se obtiene que $v(c,b) = r$ y $v(c,a) = s$. En caso contrario, 3) si $r > s > k > p$ entonces $H'' \neq H'$, y así el proceso se repite con $H' = H''$, es decir, $H''' = H'' \cup (H'' \circ H'')$. Esto es,

$$H''' = \begin{pmatrix} 1 & r & s & s \\ r & 1 & s & s \\ s & s & 1 & r \\ s & s & r & 1 \end{pmatrix}$$

Como $H''' = H''$, se obtiene la clausura transitiva después de considerar todos los posibles casos y de nuevo se cumple que $v(c,b) = r$ y $v(c,a) = s$. □

El Teorema 3 indica que, bajo ciertas condiciones, las relaciones de proximidad y similitud entre σ_m^T y φ_m^T, y entre φ_m^T y δ_m^T son equivalentes, es decir, se conservan los valores. Por lo tanto, si $r > s \geq k \geq p$ entonces $v_{t=s}(c,a) \subset v_{t=r}(c,b)$.

Teorema 4. *Considere los grupos* $G_T = \{\phi_m^T, \varphi_m^T, \sigma_m^T, \delta_m^T\}$ *y* $G_p = \{e, a, b, c\}$. *Sea* w *una relación de proximidad sobre* G_T. *Sean* $1 = w(\varphi_m^T, \varphi_m^T)$, $k = w(\varphi_m^T, \phi_m^T)$, $r = w(\varphi_m^T, \delta_m^T)$, $p = w(\varphi_m^T, \delta_m^T)$ *y* $s = w(\varphi_m^T, \sigma_m^T)$. *Si* $r > k > s \geq p$ *o* $k > r > s \geq p$, *entonces* $v(c,b) = r$ *y* $v(c,e) = k$.

Demostración. Considerando el mismo procedimiento en la determinación de la clausura transitiva de w, la prueba es análoga a la prueba del Teorema 3. □

En este caso, el Teorema 4, indica que bajo ciertas condiciones, la relaciones de similitud y proximidad entre ϕ_m^T y φ_m^T y entre φ_m^T y δ_m^T son equivalentes, es decir conservan los valores. Por lo tanto, si $r > k > s \geq p$ o $k > r > s \geq p$, entonces $v_{t=k}(c, e) \subset v_{t=r}(c, b)$.

Adicionalmente, el Teorema 3 y el Teorema 4 garantizan que es posible considerar solo los valores obtenidos por la relación de proximidad y el orden correspondiente, sin alterar la relación de orden (por inclusión) en las relaciones de equivalencia (o de los subgrupos normales). Tal resultado se emplea en el siguiente capítulo para establecer un criterio de calidad que sea invariante bajo las relaciones de similitud y de proximidad.

Una vez planteado el razonamiento descrito en esta sección es posible exponer algunas conclusiones: la relación de similitud (invariante por traslaciones) v sobre G_T (grupo formado por los grados globales de cada operador) permite formalizar un proceso de comparación entre parejas de operadores vía ordenación por inclusión. La relación de similitud se construye bajo la propuesta aquí presentada, debido a que de esta manera, se preserva la operación del grupo y los valores de pertenencia de los subgrupos borrosos de G_T. La formalización del proceso de comparación se establece en la Sección 3.2.

2.3.2. Funciones de disimilitud para los grados globales

Así como la noción de similitud aparece de forma natural para comparar elementos de un conjunto, la noción de "distancia" o en general, la noción de disimilitud, surge como concepto dual al de similitud, expresando la diferencia entre objetos, o incluso la separación o divergencia entre los mismos.

De acuerdo con lo anterior, la motivación para considerar la disimilitud entre objetos es poder integrar esta noción con su noción dual y, por ende, estudiar características de un objeto encontrando en qué medida dichas características tienen

puntos en común o difieren entre sí. Es decir, aplicando las dos nociones.

Sin embargo, la construcción propuesta en la sección anterior a partir de relaciones de similitud tiene como eje central la coincidencia entre la propiedad de *min*-transitividad de este tipo de relaciones y la definición sobre subgrupos borrosos. Esto quiere decir que el abordar el concepto de disimilitud entre objetos en el marco del grupo de operadores construido claramente no permite obtener subgrupos borrosos y, por ende, interpolar directamente lo desarrollado carece de sentido. Por lo tanto, como noción dual, se considera la disimilitud desde una perspectiva más general sin imponer el cumplimiento de la desigualdad triangular, ni la desigualdad ultramétrica. Lo anterior quiere decir que, se hará referencia a funciones de disimilitud en el sentido de lo que a partir de [86] se ha denominado cuasimétrica.

En la presente sección se construyen los elementos conceptuales para proponer un proceso de comparación vía funciones de disimilitud en el siguiente capítulo. El desarrollo propuesto nuevamente tiene como punto de partida el grupo conmutativo establecido en la Sección 2.2.

Dado un conjunto finito de objetos X con $|X| = m$ y un FCS (C, φ, ϕ, N) con n clases, se considera como antes el conjunto $G_T = \{\phi_m^T, \varphi_m^T, \sigma_m^T, \delta_m^T\}$ de los grados globales de los operadores de agregación obtenido a partir de un operador tipo promedio A, estable por negaciones fuertes. Sobre tal conjunto se establece una función de disimilitud d (de nuevo, cuasimétrica), de tal forma que se establece la siguiente matriz de disimilitud.

$$D_{d_m} = \left(\begin{array}{c|cccc} d & \phi_m^T & \sigma_m^T & \delta_m^T & \varphi_m^T \\ \hline \phi_m^T & 0 & \beta_m & \pi_m & \alpha_m \\ \sigma_m^T & \beta_m & 0 & \gamma_m & \rho_m \\ \delta_m^T & \pi_m & \gamma_m & 0 & \tau_m \\ \varphi_m^T & \alpha_m & \rho_m & \tau_m & 0 \end{array}\right)$$

Por las propiedades que cumple la función de disimilitud empleada, es inmediato que la matriz D_{d_m} es simétrica y los elementos de la diagonal principal son ceros. Por lo tanto, se prueba la Proposición 9.

Proposición 9. *Dado un conjunto X con m elementos y un FCS (C, ϕ, φ, N) con n clases. Considere el conjunto $G = \{\phi_n^T, \sigma_n^T, \delta_n^T, \varphi_n^T\}$ y una función de disimilitud usual d. Si $d(x,y) = d(N(x), N(y))$, entonces $d(\varphi_n^T, \delta_n^T) = d(\phi_n^T, \sigma_n^T)$ y $d(\delta_n^T, \sigma_n^T) = d(\phi_n^T, \varphi_n^T)$. Por lo tanto, se obtiene la siguiente matriz para la relación d,*

$$D_{\bar{d}_n} = \left(\begin{array}{c|cccc} & \phi_n^T & \sigma_n^T & \delta_n^T & \varphi_n^T \\ \hline \phi_n^T & 0 & \beta_n & \pi_n & \alpha_n \\ \sigma_n^T & \beta_n & 0 & \alpha_n & \rho_n \\ \delta_n^T & \pi_n & \alpha_n & 0 & \beta_n \\ \varphi_n^T & \alpha_n & \rho_n & \beta_n & 0 \end{array}\right)$$

donde $\bar{d}$ denota la función de disimilitud cumpliendo la propiedad específica.

Demostración. Se considera la primera igualdad, $d(\varphi_{in}^T, \delta_{in}^T) = d(\phi_{in}^T, \sigma_{in}^T)$. Sea A el operador tipo promedio (estable por negaciones fuertes N) tal que $\varphi_n^T = A\Big(\varphi_{1n}, ..., \varphi_{mn}\Big)$. A partir de la Definición 18 y la Ecuación 2.3, se tiene que para cada i con $i = 1, ..., m$, se cumple que $\varphi_{in}\Big(\mu_1(x_i), \ldots, \mu_n(x_i)\Big) = N\Big(\phi_{in}\Big(N(\mu_1(x_i)), \ldots, N(\mu_n(x_i))\Big)$

y se cumple que $\phi_{in}\Big(N\big(\mu_1(x_i)\big),\ldots,N\big(\mu_n(x_i)\big)\Big) = \sigma_{in}\Big(\mu_1(x_i),\ldots,\mu_n(x_i)\Big)$ por lo tanto, $\varphi_{in}\Big(\mu_1(x_i),\ldots,\mu_n(x_i)\Big) = N\Big(\sigma_{in}\Big(\mu_1(x_i),\ldots,\mu_n(x_i)\Big)\Big)$. Como A es estable por negaciones fuertes, entonces se cumple que

$$\varphi_n^T = A\Big(N(\sigma_{1n}),...,N(\sigma_{mn})\Big) = N\Big(A\Big(\sigma_{1n},...,\sigma_{mn}\Big)\Big) = N(\sigma_n^T)$$

De forma similar, recordando que

$$\delta_{in}\Big(\mu_1(x_i),\ldots,\mu_n(x_i)\Big) = N\Big(\phi_{in}\Big(\mu_1(x_i),\ldots,\mu_n(x_i)\Big)\Big),$$

entonces se tiene que,

$$\delta_n^T = A\Big(\delta_{1n},...,\delta_{mn}\Big) = A\Big(N\Big(\phi_{1n},...,\phi_{mn}\Big)\Big) = N\Big(A\Big(\phi_{1n},...,\phi_{mn}\Big)\Big) = N(\phi_n^T).$$

Por lo tanto, se ha comprobado que las siguientes igualdades se cumplen:

$$d(\varphi_n^T,\delta_n^T) = d(N(\sigma_n^T)\ y\ N(\phi_n^T)) = d(\sigma_n^T,\phi_n^T).$$

De forma análoga, la segunda igualdad $d(\delta_n^T,\sigma_n^T) = d(\phi_n^T,\varphi_n^T)$ se cumple, gracias a que $\delta_n^T = N(\phi_{in})$ y $\sigma_n^T = N(\varphi_n^T)$. □

El Ejemplo 8 presenta una función de disimilitud usual tal que $d(x,y) = d(N(x),N(y))$, según se ha empleado en la Proposición 9.

Ejemplo 8. *Si $N(x) = 1-x$, la función $d(x,y) = |x-y|$, cumple que $d(x,y) = d(N(x),N(y))$ para todo $x,y \in [0,1]$.*

2.3.2.1. Polinomio característico de una matriz de disimilitud

Dada una matriz cuadrada D, se asocia un polinomio que contiene información relevante sobre dicha matriz como su determinante, traza y valores propios. Formalmente se tiene la siguiente definición:

Definición 20. *Sea D una matriz cuadrada de $l \times l$. El polinomio característico de D se define como la función $f(\lambda)$ dada por $f(\lambda) = det(D - I_l\lambda)$, con I la matriz identidad $l \times l$.*

En en el contexto de la Definición 20, se resalta que la matriz $A_{\bar{d}_n}$ obtenida en la Proposición 9 se encuentra definida para cada conjunto de clases C con n clases, donde $n = 2, ..., m$. Por lo tanto, si para cada una de tales matrices se calcula su polinomio característico en la variable λ, denotado por $p_{\bar{d}_n}(\lambda)$, se obtiene que, para cada conjunto de clases C, dicho polinomio presenta los siguientes coeficientes:

$$p_{\bar{d}_n}(\lambda) = \lambda^4 - \Big(2\alpha^2 + 2\beta^2 - \pi^2 - \rho^2\Big)\lambda^2 - 4\alpha\beta\Big(\pi + \rho\Big)\lambda + \Big((\alpha+\beta)^2 - \pi\rho\Big)\Big((\alpha-\beta)^2 - \pi\rho\Big) \tag{2.5}$$

A menos de que sea necesario y sin generar ambigüedad, se escribirá α, β, ρ, π en lugar de $\alpha_n, \beta_n, \rho_n, \pi_n$.

Sobre la Ecuación 2.5 algunos resultados son inmediatos.

1. El coeficiente de λ^3 es cero, debido a que la traza de la matriz de disimilitud es cero.

2. El determinante de la matriz está dado por $\big((\alpha + \beta)^2 - \pi\rho\big)\big((\alpha - \beta)^2 - \pi\rho\big)$.

3. Las raíces del polinomio $p_{\bar{d}_n}(\lambda)$ están dadas por:

$$\lambda_{1,2} = \frac{1}{2}\pi + \frac{1}{2}\rho \pm \frac{1}{2}\sqrt{(2\alpha + 2\beta)^2 + (\pi - \rho)^2}$$
$$\lambda_{3,4} = -\frac{1}{2}\pi - \frac{1}{2}\rho \pm \frac{1}{2}\sqrt{(2\alpha + 2\beta)^2 + (\pi - \rho)^2}$$

El último resultado indica que todas las raíces del polinomio son reales. Específicamente, como α, β, ρ y π se encuentran en $[0, 1]$, se cumple el corolario 1.

Corolario 1. *Si $p_{\bar{d}_n}(\lambda)$ es el polinomio característico de la matriz $D_{\bar{d}_n}$ obtenida bajo las condiciones de la Proposición 9, entonces las raíces del polinomio $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \in [-2, 2]$.*

Demostración. Las raíces de $p_{\bar{d}_n}(\lambda)$, están dadas por

$$\lambda_i = \pm\frac{1}{2}\pi \pm \frac{1}{2}\rho \pm \frac{1}{2}\sqrt{(2\alpha + 2\beta)^2 + (\pi - \rho)^2}$$

donde $\frac{1}{2}\pi + \frac{1}{2}\rho = 1$ y $-\frac{1}{2}\pi - \frac{1}{2}\rho = -1$ si $\pi = \rho = 1$. Como $\alpha = \bar{d}(\varphi_n^T, \phi_n^T) = \bar{d}(\delta_n^T, \sigma_n^T)$, si $\alpha = 1$, entonces necesariamente $\beta = 0$ debido a que $\delta_n^T = N(\phi_n^T)$ y $\sigma_n^T = N(\varphi_n^T)$. Por lo tanto, $\frac{1}{2}\pi + \frac{1}{2}\rho + \frac{1}{2}\sqrt{(2\alpha + 2\beta)^2 + (\pi - \rho)^2} = 2$ y $-\pi\frac{1}{2} - \frac{1}{2}\rho - \frac{1}{2}\sqrt{(2\alpha + 2\beta)^2 + (\pi - \rho)^2} = -2$ □

Finalmente, se presenta un resultado de gran utilidad para el estudio de los polinomios $p_{\bar{d}_n}(\lambda)$ con $n = 2, \ldots, m$. Para ello se tiene la Definición 21.

Definición 21. *[84] Dadas dos funciones $f(z)$ y $g(z)$ de variable real o compleja. Se dice que $f(z)$ es asintóticamente equivalente o igual a $g(z)$ bajo el limite $z \to z_0$, si f y g son tales que*

$$\lim_{z \to z_0} \frac{f(z)}{g(z)} = 1$$

Lo anterior se denota por $f(z) \sim g(z)$ (cuando $z \to z_0$).

La Definición 21 establece que si dos funciones $f(z)$ y $g(z)$ son asintóticamente equivalentes (cuando $z \to z_0$), entonces el error relativo de la igualdad aproximada de $f(z)$ y $g(z)$, es decir, la magnitud

$$\frac{[f(z) - g(z)]}{g(z)}$$

es infinitamente pequeña cuando $z \to z_0$.

Claramente, la relación $f(z) \sim g(z)$ (cuando $z \to z_0$) es una relación de equivalencia. Es decir: 1) $f(z) \sim f(z)$, 2) si $f(z) \sim g(z)$ entonces,

$$\lim_{z \to z_0} \frac{g(z)}{f(z)} = \frac{1}{\lim_{z \to z_0} \frac{f(z)}{g(z)}} = 1$$

y, 3) si $f(z) \sim g(z)$ y $g(z) \sim h(z)$, entonces $f(z) \sim h(z)$.

Para el caso particular de $p_{\bar{d}_n}(\lambda)$ se cumple que, $p_{\bar{d}_{n_1}}(\lambda) \sim p_{\bar{d}_{n2}}(\lambda)$ (cuando $\lambda \to \infty$) y $p_{\bar{d}_{n_1}}(\lambda) \sim p_{\bar{d}_{n2}}(\lambda)$ (cuando $\lambda \to -\infty$) para todo $n_1, n_2 \in \{2, \dots m\}$ debido a que siempre se cumple que

$$\lim_{\lambda \to \pm\infty} \frac{p_{\bar{d}_n}(\lambda)}{p_{\bar{d}_j}(\lambda)} = 1.$$

Por lo tanto, cuando $\lambda \to \infty$ o $\lambda \to -\infty$, $p_{\bar{d}_{n_1}}(\lambda)$ y $p_{\bar{d}_{n_2}}(\lambda)$ se vuelven *esencialmente* iguales.

Los resultados presentados en esta sección permitirán establecer un proceso de comparación vía funciones de disimilitud, en el cual se estudian conjuntos de clases C con n clases, donde $n = 1, \dots, m$. Dicho proceso de comparación será presentado en la Sección 3.3.

Los conceptos y resultados presentados en el Capítulo 2 conforman las bases matemáticas sobre las cuales se platearán en el Capítulo 3, la propuesta de un proceso que permite determinar la calidad de una partición borrosa y, un índice que establece el número óptimo de clases en una partición borrosa.

Capítulo 3

Criterios de relevancia y calidad para particiones borrosas

El presente capítulo está dedicado a los siguientes propósitos. En primer lugar, se estudia el concepto de relevancia y se establece como una propiedad deseable de las clases de una partición borrosa. En segundo lugar, se busca establecer criterios que permitan, dada una partición borrosa, evaluar su calidad a partir de los elementos dados en la Sección 2.3.1 incluyendo la propiedad mencionada. Finalmente, se busca establecer un criterio que permita, dado un conjunto de objetos y un algoritmo de clasificación iterativo, determinar el número óptimo de clases. Para este último propósito se acudirá a los elementos establecidos en la Sección 2.3.2.

3.1. Sobre el concepto de relevancia

En esta sección se aborda el concepto de relevancia desde sus generalidades y elementos constitutivos. Se establecen las características que permiten definir la relevancia como una propiedad de los objetos inmersos en un contexto y se define en el marco de una partición borrosa. A partir de los sistemas de clasificación borrosa, se estudia la relevancia como una propiedad que satisfacen las clases de una partición borrosa en términos de grados.

3.1.1. Marco de referencia general

La relevancia constituye un concepto de naturaleza *vaga* y se encuentra definido por los efectos contextuales que produce, es decir, como concepto genera o valida información a partir de información ya existente. En otras palabras, la relevancia se define a partir de la existencia de un contexto.

Intuitivamente, las personas pueden distinguir información irrelevante o, en algunos casos, información más relevante de información menos relevante. Sin embargo, tal distinción no es del todo clara o no siempre está dada por un consenso generalizado. El hecho de que exista una noción en el lenguaje cotidiano de relevancia con un significado vago y variable, expone la complejidad de su definición y revela diferentes formas de abordarlo. De igual forma, las intuiciones acerca de la relevancia de un objeto son relativas (al contexto) y no hay forma de controlar exactamente qué contexto tendrá alguien en mente en un momento dado, o cómo se puede entender dicho contexto [107].

Con base en lo anterior, y considerando la definición propuesta en [107] sobre la relevancia[1], una primera idea para su estudio y aplicación en el contexto de particiones borrosas se expresa como sigue: *establecer que un elemento es relevante en un contexto dado significa que, si el objeto se agrega o elimina del contexto, entonces hay un cambio en el contexto, o hay un cambio en la información sobre dicho contexto*. En general, tener un efecto sobre el contexto es una condición necesaria para que emerja la relevancia de un objeto.

Considerando la idea anterior, se establecen entonces dos situaciones concretas en términos operacionales: determinar la relevancia de un objeto por la generación de *cambios en el contexto* o por la generación de *cambios en información del contexto*. Precisamente, sobre la última situación se profundizará en el desarrollo de la presente

[1]Relevancia: Un supuesto es relevante en un contexto si y sólo si tiene algún efecto contextual en dicho contexto

tesis.

Dado un conjunto de datos y establecida una clasificación borrosa sobre los mismos, se estudia la información acerca de la clasificación y los cambios sobre dicha información, al agregar o quitar una clase o clúster. Si bien la extracción o adición de una clase genera cambios en la clasificación, el énfasis está en la información que se puede extraer de la nueva clasificación. En este punto, es importante aclarar que extraer una clase de una clasificación puede darse en dos líneas, la primera consiste en eliminar por completo la clase dejando las clases restantes, y la segunda consiste en generar una nueva clasificación con una clase menos. Por su parte, agregar una clase en principio solo es posible generando una nueva clasificación con una clase adicional de entrada.

Ahora bien, estudiar los *cambios en información del contexto* determina dos nuevas alternativas de estudio sobre la relevancia de un objeto: 1) el estudio de los cambios en la información, una vez se ha eliminado el objeto; y 2) el estudio de la información resultante en si misma. Bajo los propósitos de este documento ambas alternativas son consideradas. La primera alternativa, expone la necesidad de comparar dos estados, *el antes* y *el después*, con el objeto y sin el objeto. La segunda alternativa, implica establecer un punto de referencia que permita determinar *el ahora*, de la información que se tiene. Por lo tanto, se proponen dos procesos que permiten llevar a cabo en términos operacionales los estudios expuestos. Por un lado, se establece un proceso de *comparación* entre la información obtenida antes y después de eliminar el objeto del contexto. Y por otro lado, con la información resultante se establece un proceso de *medición* que permita determinar el grado en que la información resultante aporta en la comprensión, caracterización o reconocimiento de estructuras subyacentes o recurrentes (patrones) del contexto.

El proceso de comparación se establece entre dos conjuntos perfectamente diferenciados, la información del contexto con el objeto y la información del contexto sin

el objeto. En este sentido, se observan entonces los cambios generados en el paso de un conjunto a otro, de un estado al otro. Por su parte, en el proceso de medición, una vez eliminado el objeto, se requiere una escala o parámetro que permita determinar qué tanto aporta, en la comprensión del contexto, la información que ha quedado.

En resumen, se propone el estudio de la relevancia en dos etapas. Una primera etapa evalúa la relevancia de un objeto comparando los cambios en la información al eliminar el objeto y otra etapa en donde se evalúa si la información resultante es relevante en el contexto.

Como concepto técnico que puede ser adecuado para ser medido por métodos computacionales, la relevancia requiere una caracterización que permita su comprensión formal para el uso computacional. Por lo tanto, se propone un nuevo enfoque o comprensión sobre el concepto de relevancia en el marco borrosos y los medios para evaluarla y medirla, particularmente, con referencia al estudio de particiones borrosas y las clases, agrupamientos o clúster obtenidos.

De acuerdo con lo anterior, la nueva propuesta funciona bajo las siguientes consideraciones: el contexto hace alusión a la clasificación obtenida sobre un conjunto de datos; y la información del contexto corresponde a la información proporcionada por la evaluación de dicha clasificación a través de la aplicación de las reglas recursivas y sus negaciones, según la tripleta de De Morgan seleccionada. Esto es, con respecto a una partición borrosa comparamos los grados de redundancia, cobertura, no relevancia y no cobertura del conjunto de clases que se estudian. Como se mencionó anteriormente, la relevancia se representa aquí como un concepto vago, borroso y, por lo tanto, su estudio es una cuestión de grados.

A continuación se plantea en términos formales la propuesta y se profundiza en los detalles y consideraciones necesarias para su comprensión y aplicación.

3.1.2. Relevancia de una clase en particiones borrosas

Considerando un FCS, la relevancia de una clase en una partición borrosa ha sido abordada en [4] mediante una metodología basada en contraste de hipótesis. En dicha propuesta, la relevancia de una clase se evalúa a través del operador disyuntivo φ, comparando los grados de cobertura de toda la partición con los obtenidos sin la clase en estudio. De manera similar, en [29] se realiza un primer acercamiento al estudio de la relevancia de una clase con el operador disyuntivo φ, esta vez sin usar una prueba estadística. En ambas propuestas, se considera una sola regla recursiva. En general, una gran cantidad de propuestas para determinar la calidad de una partición difusa se basan en el estudio de un único índice o la medición de una sola propiedad. Bajo el enfoque propuesto en esta tesis, se consideran todas las reglas recursivas dadas por el sistema de clasificación borroso y sus negaciones, lo cual implica una consideración de varios índices.

Según se ha planteado anteriormente, bajo nuestro enfoque, la relevancia es entendida como la necesidad de incluir, excluir o no excluir, una clase o familia de clases de una partición borrosa dada. Alguna de dichas necesidades puede plantearse porque, por ejemplo, una clase puede estar generando altos grados de solapamiento sin mejorar la cobertura, o porque, aunque una clase genera altos grados de cobertura, algunos elementos del conjunto de datos, se explican mejor por otra clase.

Por lo tanto, se considera la relevancia, como una propiedad que permite estudiar el problema de la reducción de la dimensionalidad en una partición borrosa. Esto es, dado un conjunto de datos, se pretende encontrar la mejor partición borrosa con el menor número de clases. El Ejemplo 9 ilustra este problema.

Ejemplo 9. *Sea* $C = \{c_1, c_2, c_3, c_4\}$ *sobre el conjunto de datos* $X = \{x_1, x_2, x_3, x_4, x_5\}$. *La Tabla 3.1 presenta los grados de pertenencia de los cinco datos dentro de las cuatro clases de* C. *Se ha seleccionado el FCS dado por*

- $\varphi_4(\mu_1(x), \mu_2(x), \mu_3(x), \mu_4(x)) = \max(\mu_1(x), \mu_2(x), \mu_3(x), \mu_4(x))$,

- $\phi_4(\mu_1(x), \mu_2(x), \mu_3(x), \mu_4(x)) = \min(\mu_1(x), \mu_2(x), \mu_3(x), \mu_4(x))$
- $N(\mu(x)) = 1 - \mu(x)$.

Las dos últimas columnas en la Tabla 3.1 presentan los grados de cubrimiento y solapamiento de la partición C para cada objeto $x_i \in X$. El grado de cubrimiento global y redundancia global se obtienen usando el operador $A = \frac{1}{m}\sum_{k=1}^{m}\theta_{kn}$ como se muestra en la última fila de la tabla.

Objetos	c_1	c_2	c_3	c_4	φ_4	ϕ_4
x_1	0.3	0.4	0.2	0.1	0.4	0.1
x_2	0.6	0.2	0.12	0.08	0.6	0.08
x_3	0.8	0.1	0.07	0.03	0.8	0.03
x_4	0.2	0.5	0.25	0.05	0.5	0.05
x_5	0.1	0.1	0.79	0.01	0.79	0.01
					$\varphi_4^T = 0{,}618$	$\phi_4^T = 0{,}054$

Tabla 3.1: *Grado de cubrimiento y solapamiento global*

Sin embargo, se observa que la clase c_4 puede ser eliminada sin deteriorar el grado de cubrimiento y el grado de solapamiento de las clases restantes. Adicionalmente, después de eliminar dichas clases se obtienen los siguientes valores para los grados globales, $\varphi_3^T = 0{,}618$ y $\phi_3^T = 0{,}138$. De acuerdo con la situación planteada, surgen de forma natural las siguientes preguntas: ¿Cómo determinar si una clase se puede eliminar? ¿Es mejor eliminar una clase o volver a realizar la clasificación considerando una clase menos desde el inicio? Si se elimina una clase y el grado de solapamiento aumenta, ¿Hasta que punto es aceptable incrementar el solapamiento de las clases restantes?

Por lo tanto, la propiedad de relevancia, en el caso de particiones borrosas, es un elemento clave por cuanto un objeto puede pertenecer simultáneamente a varias clases y, en ocasiones, su pertenencia a una de estas clases puede no proporcionar información discriminante sobre el objeto.

Considerando los anteriores interrogantes, se propone estudiar e integrar toda la información disponible por el FCS (reglas recursivas y sus negaciones) y establecer un procedimiento que no considere las reglas recursivas de forma aislada. Un primer acercamiento a esta idea se puede encontrar en [27]. A partir de tal propuesta, emanan dos consideraciones claves: en primer lugar, ¿cómo establecer una estructura que permita considerar, integrar y comparar las reglas recursivas y sus correspondientes negaciones? En segundo lugar, una vez que se logra establecer una estructura consistente, ¿cómo determinar un criterio que permita determinar si se deben eliminar o agregar clases? Es decir, ¿cómo encontrar el número óptimo de clases o clústeres del conjunto de datos, considerando las propiedades de cubrimiento y redundancia del conjunto de clases y la relevancia de cada clase?

Lo anterior implica la necesidad de estudiar la partición borrosa desde dos perspectivas: por un lado, medir las propiedades de toda la partición o de un subconjunto de clases de la partición (propiedades globales); y por otro lado, medir propiedades de cada clase (propiedades locales).

La propuesta de un conjunto de criterios para evaluar una partición borrosa dada se establece entonces bajo el siguiente razonamiento: considerando los resultados presentado en la Sección 2.3.1, se considera un sistema de clasificación borroso (C, φ, ϕ, N), donde $C = \{c_1, \ldots, c_n\}$ describe lo que en esta sección se ha denominado como el contexto. Por su parte, φ, ϕ, N proporcionan información sobre ese contexto. Se plantea la estructura de grupo conmutativo formado por lo grados globales de las reglas recursivas y sus negaciones, es decir, $G_T = \{\varphi_m^T, \phi_m^T, \sigma_m^T, \delta_m^T\}$, en concordancia con la Definición 19 y cumpliendo la Proposición 2 y la Proposición 3. Para comparar los elementos de dicho grupo, se emplea una relación de similitud invariante por traslaciones v sobre G_T de acuerdo con la Ecuación 2.3 cumpliendo las Proposiciones 7 y 8 y los Teoremas 1 y 2. Dicha relación v se obtiene a partir de una relación de proximidad w cumpliendo los Teoremas 3 y 4.

A partir de las clases de equivalencia correspondientes v_t obtenidas a partir de v como se describe en el Teorema 1, se establece un proceso de comparación considerando la inclusión de clases. Con base en dicha comparación, se establece un criterio que permite determinar cuándo una partición tiene niveles óptimos de cubrimiento y redundancia para, posteriormente, determinar si hay clases o conjuntos de clases irrelevantes.

3.2. Criterios de calidad y relevancia para una partición borrosa

En general, una partición es de calidad cuando presenta un alto grado de cubrimiento y un bajo grado de solapamiento. Por lo tanto, cuando se consideran los grados de no cubrimiento y de no solapamiento, equivalentemente una partición se considera de calidad si tiene un bajo grado de no cubrimiento y un alto grado de no solapamiento.

Con el fin de integrar lo expuesto en el párrafo anterior, se define en la presente sección, un criterio de calidad para una partición borrosa y un criterio de relevancia para las clases de dicha partición. Tales criterios surgen a partir de un proceso de comparación que es descrito a través de un algoritmo desarrollado para tal fin. Dicho algoritmo permite, a través de un proceso iterativo, evaluar si una partición borrosa es de *calidad* y si todas sus clases son relevantes, lo cual determina lo que se ha definido como una partición de *alta calidad*. Adicionalmente se define un criterio que permite comparar dos particiones borrosas que cumplen con el criterio de calidad y el criterio de relevancia, estableciendo por lo tanto, un primer acercamiento a encontrar el número óptimo de clases. Parte de las ideas expuestas en la presente sección se encuentran ya publicadas en [24].

En la Sección 2.3.1 se estableció una interpretación sobre los resultados obtenidos

con la aplicación de la relación de proximidad w sobre G_T. Con el fin de ampliar la comprensión de los criterios a definir, se recuerda tal interpretación: en general $w(x, y)$ se interpreta como el grado en que x y y son *aproximadamente iguales*. Por lo tanto, si φ_n^T representa el grado de cubrimiento de un conjunto de n clases para m objetos (grado global) y ϕ_n^T el grado de solapamiento, entonces el grado en que ambas propiedades son *aproximadamente iguales*, se puede ver como aquella parte en que ambas coinciden. En ultimas, de acuerdo con la intuición mas general sobre la similitud (ver [73]), tal característica de un par de objetos está directamente asociada con la información que ambos objetos tienen en común y con la información que define cada objeto.

De acuerdo con lo anterior, se interpretan entonces los valores obtenidos por la relación w sobre G_T de la siguiente forma: $w(\varphi_m^T, \phi_m^T)$ proporciona el grado que cubre la redundancia de las clases, $w(\varphi_m^T, \delta_m^T)$ proporciona el grado de cubrimiento sin considerar la redundancia de las clases y, $w(\varphi_m^T, \sigma_m^T)$ proporciona el grado en que el cubrimiento y el no cubrimiento son aproximadamente iguales. Por lo tanto, es deseable, por ejemplo, que el grado que cubre la redundancia de las clases, sea bajo y menor que el grado que cubren las clases sin solapamiento.

Se propone entonces el primer criterio de *calidad*. Sea X un conjunto de objetos con $|X| = m$, (C, φ, ϕ, N) un FCS con $C = \{c_1, ..., c_n\}$, donde C es obtenido a partir de un algoritmo de clasificación borroso iterativo y no jerárquico, por ejemplo, a través de fuzzy c-means o possibilistic c-means. Considere los grupos $G_T = \{\phi_m^T, \varphi_m^T, \sigma_m^T, \delta_m^T\}$ y $G_p = \{e, a, b, c\}$ junto a una relación de similitud invariante por traslaciones v sobre G_p. Por lo tanto, es posible calcular las relaciones de equivalencia $v_t = \{(x, y)|v(x, y) \geq t\}$ para todo $t \in [0, 1]$. Bajo esta perspectiva, si $0 \leq t_1 < \cdots < t_p = 1$, entonces $v_{t_p} \subseteq \cdots \subseteq v_{t_1} = G_T \times G_T$.

Sean $1 = w(\varphi_m^T, \varphi_m^T)$, $k = w(\varphi_m^T, \phi_m^T)$, $r = w(\varphi_m^T, \delta_m^T)$ y $s = w(\varphi_m^T, \sigma_m^T)$ y sean $x, y \in G_p = \{e, a, b, c\}$. Entonces se definen la relaciones de equivalencia

$K = v_{t=k}(x, y) = \{(x, y)|v(x, y) \geq k\}$,

$R = v_{t=r}(x, y) = \{(x, y)|v(x, y) \geq r\}$,

$S = v_{t=s}(x, y) = \{(x, y)|v(x, y) \geq s\}$.

Considerando los resultados de los Teoremas 3 y 4, se plantea el siguiente criterio:

Criterio de calidad. *Una partición borrosa C con n clases es una partición de calidad si $r > s > k$, o de forma equivalente, si $S \subset R$. En caso contrario, se dice que la partición es de baja calidad. Por lo tanto, el valor de r determina el grado de calidad de la partición borrosa, y cuanto mayor sea, mejor.*

Es importante recordar que r proporciona el grado de cubrimiento sin considerar el solapamiento de las clases.

En la misma línea, se puede extender el argumento empleado hasta ahora para un criterio de *relevancia* sobre cada una de las clases de la partición borrosa.

Criterio de relevancia. *Dada una partición borrosa C con n clases, sea $C_i \subseteq C = \{c_1, \ldots, c_n\}$ una familia de clases no vacías fija, tal que C_i es el subconjunto de $n-1$ clases obtenidas al eliminar la clase $c_i \in C$, con $i = \{1, ..., n\}$. Sean k_i, r_i, y s_i los correspondientes valores para w sobre C_i. Si $k_i > s_i$, o de forma equivalente se cumple que $S_i \subset K_i$, entonces c_i es una clase de alta relevancia para la partición C. En caso contrario, se dice que c_i es una clase de baja relevancia para la partición C.*

El criterio procede eliminando de forma iterativa, una clase del conjunto de clases en consideración, siendo esta clase cada vez distinta, para estudiar su relevancia. En este sentido, la clase eliminada será relevante si la agregación de las clases restantes indica que el grado de *cubrimiento del solapamiento* es mayor que el grado de *cubrimiento sin considerar el solapamiento*. En otras palabras, una clase se considera relevante si al ser eliminada de C, entonces disminuye el grado de cubrimiento en una proporción mayor que el grado de solapamiento.

Por lo tanto, podemos establecer en términos más formales la intuición básica que respalda ambos criterios, a través de la la Definición 22.

Definición 22. *Una partición borrosa es llamada de alta calidad, si todas sus clases son relevantes.*

De acuerdo con los criterios establecidos, es posible desarrollar un procedimiento para evaluar la calidad de una partición borrosa dada. Tal procedimiento es sintetizado en el Algoritmo 1.

Dicho algoritmo procede de la siguiente manera: en el primer paso, una partición $C = \{c_1, c_2\}$ es calculada a través de un algoritmo de clasificación borrosos iterativo no jerárquico. Para efectos prácticos y debido a su uso generalizado y a los óptimos resultados obtenidos en diversas aplicaciones, se propone emplear el algoritmo fuzzy c-means. A partir de la relación de proximidad w, el Algoritmo 1 determina si la partición C es de calidad. En este punto es importante anotar que en el caso de $n = 2$ (dos clases), la relevancia de las clases no es evaluada debido a que bajo la propuesta planteada, la agregación de una sola clase no es considerada. Por lo tanto, si la partición con dos clases es de calidad, entonces el algoritmo finaliza. En caso contrario, el algoritmo incrementa el número de clases en una y retorna al paso 1 hasta encontrar una partición de alta calidad. Es decir, una partición en la cual todas sus clases son relevantes.

Algorithm 1 Determinación de una partición borrosa de calidad

Input: X un conjunto finito de objetos, $|X| = m$, reglas recursivas φ y ϕ y la negación N (correspondientes a un FCS), una relación de proximidad w y $n = 2$. Sea z un entero positivo denotando el máximo número de clases a considerar.

Proceso
Global
while $r < s \leq k$ **do**
  if $n < z$ **then**
    $n = n + 1$,
    Global.
  else if $n = z$ **then**
    RETURN "Partición de calidad no encontrada" and END.
if $r > s > k$ **and** $c = 2$ **then**
  RETURN C and END.
else if $r > s > k$ and $c > 2$ **then**
  Relevancia
while $s_i \leq k_i$ **do**
  if $n < z$ **then**
    $n = n + 1$,
    Global,
    Relevancia
  else if $k_i > s_i$ **then**
    RETURN C and END.
  else if $n = z$ **then**
    RETURN "Partición de alta calidad no encontrada" and END.
End Proceso

SubProceso *Global*
  Calcule C para n ▷ n clases obtenidas con algoritmo fuzzy c-means.
  Calcule $\phi_m^T, \varphi_m^T, \sigma_m^T$ y δ_m^T y considere $G_T = \{\phi_m^T, \varphi_m^T, \sigma_m^T, \delta_m^T\}$
  Calcule w sobre G_T. Sean $k = w(\varphi_n^T, \phi_n^T)$, $r = w(\varphi_n^T, \delta_n^T)$, $s = w(\varphi_n^T, \sigma_n^T)$.
Fin SubProceso

SubProceso *Relevancia*
for $i = 1 : n$ **do**
  Calcule C_i ▷ $n - 1$ clases obtenidas al eliminar la clase i.
  Calcule $\phi_i^T, \varphi_i^T, \sigma_i^T$ y δ_i^T para cada C_i y considere $G_i = \{\phi_i^T, \varphi_i^T, \sigma_i^T, \delta_i^T\}$
  for cada C_i **do**
    calcule w sobre G_i
Fin SubProceso

Se observa del algoritmo propuesto que el proceso finaliza cuando una partición de alta calidad ha sido encontrada, es decir, una partición cumpliendo los criterios

de calidad y relevancia y a su vez, siendo la partición con el menor número de clases que satisfacen dichos criterios. Sin embargo, puede ocurrir que no exista un óptimo número de clases. En este caso se sugiere reiniciar el algoritmo fuzzy c-means con una semilla diferente o con un parámetro de borrosidad diferente.

De igual forma, puede ocurrir que existan dos particiones cumpliendo todos los criterios propuestos. En principio, la partición con el menor número de clases es la deseada, sin embargo, esto no garantiza que no se pierda o no esté siendo considerada información relevante. Por lo tanto, se propone un criterio adicional para comparar dos particiones de alta calidad cuando el Algoritmo 1 es usado varias veces bajo diferentes configuraciones (semillas o valores de parámetros borrosos). Es importante resaltar que el criterio establecido para tal fin, no es usado en el Algoritmo 1.

Criterio de comparación. *Sea C_{n_1} y C_{n_2} dos particiones borrosas de alta calidad sobre X. Sea $r_1 = w(\varphi^T_{n_1}, \delta^T_{n_1})$ y $r_2 = w(\varphi^T_{n_2}, \delta^T_{n_2})$ los valores de w para cada partición, respectivamente. Entonces C_{n_1} es mejor que C_{n_2} si $r_1 > r_2$.*

El criterio de comparación establece que frente a dos particiones que han sido evaluadas como de *alta calidad*, la que mejores resultados presenta es aquella que tiene el más alto grado de cubrimiento sin considerar la redundancia de las clases.

3.2.1. Operadores de agregación para medir calidad y relevancia

Los criterios de relevancia y calidad para evaluar particiones borrosas planteados en la sección anterior, establecen como elemento primordial para su funcionamiento, la comparación de los valores dados por $k = w(\varphi^T_m, \phi^T_m)$ y $s = w(\varphi^T_m, \sigma^T_m)$, lo cual implica que la existencia de una igualdad entre k y s dependerá exclusivamente de la igualdad entre los valores obtenidos a través de los operadores ϕ^T_m y σ^T_m. En este sentido, una pregunta que surge de forma natural es, ¿existen algunos casos en los

cuales la igualdad de estos dos valores depende de los operadores empleados y no de una característica propia y única de la partición evaluada?

Así como en la Sección 2.2 ya se ha establecido que el operador ϕ (operador conjuntivo) no puede ser *autodual*, la anterior pregunta busca excluir aquellos operadores para los cuales la igualdad entre los grados de solapamiento y no cubrimiento de una partición estudiada, no sea una característica de la partición estudiada sino una propiedad de los operadores empleados.

Retomando lo planteado en [105] (incluso ver [21, 45]), para el caso de los operadores autoduales se satisface la ecuación funcional $\phi_2(x,y) + \phi_2(1-x, 1-y) = 1$, lo cual para la propuesta aquí planteada corresponde a la igualdad entre $\phi_2(x,y)$ y $\varphi_2(x,y)$, para $N = 1 - x$. Ahora bien, si lo que se desea estudiar corresponde con la igualdad entre $\phi_2(x,y)$ y $\sigma_2(x,y)$, entonces se indaga por el cumplimiento de la ecuación funcional

$$\phi_2(x,y) - \phi_2(N(x), N(y)) = 0$$

con N una negación fuerte, o de forma equivalente, $\phi_2(x,y) = \sigma_2(x,y)$. La anterior ecuación se puede interpretar como una versión *débil* de la ecuación abordada en [105]. Cabe resaltar que los resultados de esta sección se encuentran publicados en [26].

Con base en el anterior razonamiento, se establece la Proposición 10.

Proposición 10. *Dado un sistema de clasificación borrosa (C, ϕ, φ, N) con $N = 1 - \mu(x)$ y ϕ un operador simétrico, entonces se cumple que*

$$\phi_2(\mu_1(x), \mu_2(x)) - \phi_2(N(\mu_1(x)), N(\mu_2(x))) = 0$$

si y solo si $\mu_1(x) + \mu_2(x) = 1$.

Demostración. (Necesidad) Como $\phi_2(\mu_1(x), \mu_2(x)) = \phi_2(N(\mu_1(x)), N(\mu_2(x)))$, si se

supone cierto que $\mu_1(x) + \mu_2(x) \neq 1$, entonces al considerar $\mu_1(x) = 0$ se tiene que $\mu_2(x) \neq 1$, y por lo tanto, $0 = \phi_2(0, \mu_2(x)) = \phi_2(1, 1 - \mu_1(x))$, lo cual no es posible. Así pues es necesario que $\mu_1(x) + \mu_2(x) = 1$.

(Suficiencia) Si $\mu_1(x)+\mu_2(x) = 1$ entonces $\mu_1(x) = 1-\mu_2(x)$ y como ϕ es simétrico, entonces se cumple que $\phi_2(\mu_1(x), 1 - \mu_1(x)) = \phi_2(1 - \mu_1(x), \mu_1(x))$, y por lo tanto, $\phi_2(\mu_1(x), 1-\mu_1(x)) = 1-\varphi_2(\mu_1(x), 1-\mu_1(x))$ y así $\phi_2(\mu_1(x), 1-\mu_1(x)) = \sigma_2(\mu_1(x), 1-\mu_1(x))$ o, de forma equivalente, $\phi_2(\mu_1(x), \mu_2(x)) - \phi_2(N(\mu_1(x)), N(\mu_2(x))) = 0$.

□

La Proposición 10 establece que para cualquier tripleta de Morgan (ϕ, φ, N) tal que ϕ_2 sea simétrico y $N = 1-\mu(x)$, entonces la ecuación funcional $\phi_2(\mu_1(x), \mu_2(x))-\phi_2(N(\mu_1(x)), N(\mu_2(x))) = 0$ se satisface.

Con base en lo anterior, y siguiendo la notación de la Sección 1.1, se plantea la siguiente definición

Definición 23. *Un operador de agregación $A_2 : [0,1]^2 \to [0,1]$ se denomina débilmente autodual si $\forall x, y \in [0,1]$ tal que $x + y = 1$, entonces $A(x, y) = A(N(x), N(y))$ con N una negación fuerte.*

Existe un amplio conjunto de operadores de agregación que cumplen la propiedad de ser *débilmente autoduales*. En general, como se puede deducir fácilmente de la Proposición 10, todos los operadores bidimensionales simétricos junto con la negación fuerte $N(x) = 1 - x$ cumplen la propiedad. El Ejemplo 10 presenta operadores autoduales y no autoduales.

Ejemplo 10. *Algunos operadores que cumplen la propiedad de ser auto duales con la negación $N = 1 - x$*

- $A(x, y) = \text{mín}\{x, y\}$
- $A(x, y) = xy$

- $A(x,y) = \frac{3xy}{1+2xy}$

Algunos operadores que no cumplen la propiedad de ser autoduales

- $A(x,y) = xy^2$ *con* $N = 1 - x$
- $A(x,y) = \frac{xy}{\frac{2}{3}+\frac{1}{3}(x+y+xy)}$ *con* $N = \frac{1-x}{1+2x}$

Una vez se ha establecido la propiedad descrita en la Definición 23, se procede a determinar el conjunto de operadores que pueden emplearse en los criterios presentando en la Sección 3.2, junto con el algoritmo propuesto. En el caso que se desee evaluar una partición borrosa en el sentido de Ruspini [100], es necesarios que el operador de agregación ϕ_n satisfaga las siguientes condiciones:

- $\phi_n(x_1, ..., x_n)$ debe ser un operador conjuntivo en el sentido de que $\phi_n(x_1, ..., x_n) = 0$, si existe algún $x_i = 0$.
- $\phi_n(x_1, ..., x_n)$ *no* debe ser autodual o, equivalentemente, no debe ser estable por negaciones fuertes (ver Definición 10).
- $\phi_2(x_1, x_2)$ no debe ser débilmente autodual.

Bajo las anteriores condiciones, es claro que un operador débilmente autodual no genera una evaluación consistente en el caso de una partición de Ruspini con dos clases, por lo cual, el resultado para tal partición siempre será negativa, es decir, el Algoritmo 1 siempre descartará dicha partición como una partición de calidad.

Por otro lado, si la partición que se desea evaluar no cumple las características de una partición de Ruspini, entonces no es necesaria la imposición referida a que $\phi_n(x_1, ..., x_n)$ no sea débilmente autodual.

De acuerdo con lo anterior, a través de la presente sección se definió una propiedad que cumplen algunos operadores de agregación bidimensionales y la cual ha sido denominada como la propiedad de ser *débilmente autodual*. Dicha propiedad relaciona, en

el caso particular de los FCS, los grados de solapamiento y los grados de no cubrimiento de una partición borrosa. En este contexto se ha determinado que los operadores empleados para evaluar la calidad de una partición borrosa, bajo lo propuesto en la sección 3.2, no deben cumplir dicha propiedad para hacer de la evaluación, un proceso consistente.

3.3. Número óptimo de clases de una partición borrosa

El proceso de comparación sintetizado en el Algoritmo 1 a partir de relaciones de similitud sobre G_T permite adicionalmente evaluar si una partición específica, para cualquier n con $n \in \{1, \ldots, m\}$, es de calidad o de alta calidad: para ello basta con cambiar la entrada del algoritmo de $n = 2$ a cualquier n deseado. De igual forma, permite a partir de un conjunto de objetos X determinar la primera partición borrosa que es de calidad, considerando de forma iterativa valores desde $n = 2$ hasta máximo $n = m$. Sin embargo, tan pronto el algoritmo ha encontrado la primera partición de alta calidad, se detiene.

Un aspecto que adquiere relevancia tanto en la aplicación como en la evaluación de particiones borrosas, es poder comparar dos particiones que en general, presenten propiedades deseables de forma simultanea. Al finalizar la Sección 3.2, precisamente se estableció un criterio que permite en una primera fase un proceso de comparación en este contexto. Sin embargo, dicho criterio considera apenas una propiedad sobre el conjunto de clases; el grado de *cubrimiento sin solapamiento*. Aunque fundamental, este criterio deja de lado otras propiedades y, por ende, información útil para el tomador de decisiones.

Adicionalmente, el Algoritmo 1 no está diseñado para considerar particiones con un número de clases mayor a la primera partición de *alta calidad*. Por lo tanto, a menos que exista una variación en los parámetros del algoritmo, no es posible determinar

si existe otra partición que cumpla con los criterios expuestos. De acuerdo con lo anterior, y como es natural en la evaluación de una clasificación sobre un conjunto de objetos, es oportuno complementar el proceso hasta ahora desarrollado con un proceso de comparación de naturaleza externa, es decir, con algún tipo de referencia externa o umbral sobre el cual no se está dispuesto a sobrepasar. En este sentido, establecer un umbral implica tener la posibilidad de considerar todos los elementos dentro del umbral. En el contexto de la clasificación borrosa esto significa que dado un conjunto de objetos y un conjunto de particiones borrosas sobre dichos objetos, se quiere determinar cuál es la mejor partición, es decir, cuál es el número de clases que mejor clasifican el conjunto de objetos, o en otras palabras, cuál el número óptimo de clases.

Teniendo en cuenta lo anterior, se propone entonces un proceso de comparación de carácter externo en el cual se establece una situación ideal o umbral. En este caso, es natural considerar dicho umbral a través del planteamiento de una partición crisp o nítida sobre el conjunto de objetos. Esta situación se determina como ideal teniendo en cuenta que si clasifica un conjunto de objetos de naturaleza borrosa, lo mejor que puede ocurrir es que tales objetos queden completamente representados por alguna clase.

3.3.1. Optimización con polinomios característicos

Considerando los elementos desarrollados en la Sección 2.3.2 sobre funciones de disimilitud, adquiere sentido preguntarse ¿en qué medida la partición considerada está alejada del umbral? En esta perspectiva, si se comparan los grados globales de una partición nítida a través de una función de disimilitud, bajo las condiciones de la Proposición 9, entonces se obtiene la matriz D_c presentada a continuación.

$$D_c = \left(\begin{array}{c|cccc} & \phi_m^T & \sigma_m^T & \delta_m^T & \varphi_m^T \\ \hline \phi_m^T & 0 & 0 & 1 & 1 \\ \sigma_m^T & 0 & 0 & 1 & 1 \\ \delta_m^T & 1 & 1 & 0 & 0 \\ \varphi_m^T & 1 & 1 & 0 & 0 \end{array}\right)$$

Naturalmente, se está considerando una partición en la cual el grado de solapamiento es igual a cero y el grado global de cubrimiento es igual a uno. Por lo tanto, para dicha matriz de disimilitud se tiene el polinomio característico dado en la Ecuación 3.1.

$$p_c(\lambda) = \lambda^4 - 4\lambda^2. \tag{3.1}$$

Se recuerda que, en general, el polinomio característico de una matriz de disimilitud sobre el grupo G_T para un FCS (C, φ, ϕ, N) con n clases y $|X| = m$ está dado por la Ecuación 2.5 y ha sido denotado por $p_{\bar{d}_n}$. Por lo tanto, considerando $p_c(\lambda)$ y $p_{\bar{d}_n}(\lambda)$, se propone el problema de optimización descrito en la Ecuación (3.2).

$$\arg\min_n \left\{ \int_{-a}^{a} \left| p_c(\lambda) - p_{\bar{d}_n}(\lambda) \right| d\lambda \right\} \tag{3.2}$$

con $n = 2, ..., k$, $k \leq m$ y $a \geq 2$.

Para cada n es inmediato que la integral impropia

$$\int_{-\infty}^{\infty} \left| p_c(\lambda) - p_{\bar{d}_n}(\lambda) \right| d\lambda$$

es divergente. En la misma línea, se verifica fácilmente que

$$p_c(\lambda) \sim p_{\bar{d}_n}(\lambda) \text{ cuando } \lambda \to \pm\infty$$

es decir, la función dada por $p_{\bar{d}_n}(\lambda)$ para cada n y la función dada por $p_c(\lambda)$ son asintóticamente equivalentes (cuando $\lambda \to \pm\infty$) y, por lo tanto, el error relativo de la igualdad aproximada entre cualesquiera de ellas es infinitamente pequeña. La funciones son aproximadamente iguales cuando $\lambda \to \pm\infty$. Adicionalmente, por el Corolario 1, se sabe que todas las raíces de dichos polinomios son reales y se encuentran en $[-2, 2]$, en particular $\lambda_1 = -2$ y $\lambda_1 = 2$ son la raíces de $p_c(\lambda)$. Estos argumentos son la base para considerar la integral propuesta en la Ecuación (3.2) con los limites establecidos en busca de determinar la distancia entre las curvas, aunque la integral impropia sea divergente.

Por lo tanto, el problema planteado en la Ecuación (3.2) se interpreta como sigue: encontrar el conjunto de clases C_n con n clases tal que su polinomio característico esté lo más *cerca* posible al polinomio característico del umbral o caso ideal.

En particular, si se considera un conjunto de particiones borrosas $\mathbb{C}_p = \{C_2, ..., C_k\}$, donde $p = k - 1$ es el numero de particiones obtenidas a través de un algoritmo de clasificación borroso iterativo y, cada partición C_n tiene n clases, entonces se propone la Definición 24 la cual se encuentra publicada en [23].

Definición 24. *Dado un conjunto X con $|X| = m$ y dado el conjunto de particiones borrosas $\mathbb{C}_p = \{C_2, ..., C_k\}$, con $k < m$. Sea (C_n, ϕ, φ, N) un FCS con ϕ, φ, N fijos para cada C_n y $n = 2, ..., k$. El número óptimo de clases borrosas para X, denotado por n_o, está dado por,*

$$n_o = \arg\min_n \left\{ \int_{-a}^{a} \left| p_c(\lambda) - p_{\bar{d}_n}(\lambda) \right| d\lambda \right\}$$

con $a \geq 2$.

Bajo las condiciones de la Definición 24 se considera un conjunto $\mathbb{C}_p$ de p particiones donde $p = k - 1$ y tal conjunto está acotado por $m - 1$ particiones. De igual forma, la definición plantea que el conjunto $\mathbb{C}_p$ contiene la partición inicial con $n = 2$ clases y cada partición adicional, contiene una clase adicional, es decir, el conjunto está ordenado por el número de clases.

Los resultados presentados en el Capítulo 3 conforman la propuesta para la evaluación de la calidad y la determinación del número óptimo de clases de una partición borrosa. Dicha propuesta tiene como eje fundamental el uso de sistemas de clasificación borrosa, una estructura de grupo conmutativo sobre dichos sistemas y, el uso de relaciones de similitud y disimilitud para comparar los elementos del grupo. En el capítulo 4 se aplica la propuesta al procesamiento de imágenes y se establecen los resultados de tal aplicación.

Capítulo 4

Aplicaciones al procesamiento de imágenes

El presente capítulo tiene como objetivo aplicar tanto los criterios de calidad y relevancia propuestos en la Sección 3.2, así como la definición sobre el número óptimo de clases propuesto en la Sección 3.3, en el problema específico de segmentación de imágenes, es decir, considerar el problema de clasificación no supervisado de obtener clases de píxeles similares.

En la primera parte, se detalla la aplicación del algoritmo propuesto a un caso particular con una única imagen. En la siguiente sección, se aplica el cálculo del número óptimo de clases, igualmente para un caso particular. Finalmente, en la última sección de este capítulo se aplican de forma simultanea los criterios, la definición de número óptimo y una serie de índices comúnmente usados para la validación de agrupamientos, a la imágenes de las secciones anteriores y a una tomografía axial computarizada y se comparan los resultados obtenidos. Las imágenes tratadas son de interés debido a la presencia de bordes y zonas poco nítidas o, sin límites claramente definidos.

4.1. Aplicación de los criterios de calidad y relevancia

Con el fin de aplicar los criterios definidos y el Algoritmo 1 propuesto en la Sección 3.2, se ha seleccionado la imagen presentada en la Figura 4.1.

Figura 4.1: *Aurora boreal*

Se considera la tripleta recursiva definida por,

- Sistema de clasificación borrosa (FCS) conformado por:

  1. $\phi_n(\mu_1(x), \ldots, \mu_n(x)) = \prod_{i=1}^{n} \mu_i(x)$
  2. $\varphi_n(\mu_1(x), \ldots, \mu_n(x)) = \dfrac{\prod_{i=1}^{n}\left(1+2(\mu_i(x))\right) - \prod_{i=1}^{n}(1-\mu_i(x))}{\prod_{i=1}^{n}\left(1+2(\mu_i(x))\right) + \prod_{i=1}^{n}(1-\mu_i(x))}$
  3. $N(\mu(x)) = \frac{1-x}{1+2x}$

- Algoritmo de clasificación iterativo no jerárquico: Fuzzy c-means
- Función de proximidad: $w(\theta_m^T, \lambda_m^T) = 1 - |\theta_m^T - \lambda_m^T|$.
- Número máximo de clases a considerar: $z = 8$

Se emplea la relación de proximidad sobre G_T dada por $w(\theta_m^T, \lambda_m^T) = 1 - |\theta_m^T - \lambda_m^T|$.

Figura 4.2: *Clases obtenidas por la aplicación del algoritmo fuzzy c-means para* $c = 2$ *en aurora boreal*
(izquierda: clase 1, derecha: clase 2). La escala de grises representa los grados de pertenencia de cada píxel a cada clases, donde negro= 0 y blanco =1..

De acuerdo con el Algoritmo 1 se inicia calculando una partición con $n = 2$ a través del algoritmo fuzzy c-means. Posteriormente, empleando la relación w se obtienen los valores $r = 0{,}97$, $s = 0{,}14$, $k = 0{,}17$. Para el caso particular, como $s < k$ entonces la partición obtenida es una partición de *baja calidad.* Las clases correspondientes a la partición sobre la imagen son presentadas en la Figura 4.2.

A partir de la visualización de la segmentación obtenida, se puede inferir que la baja calidad de la partición se evidencia en algunas regiones donde los objetos con diferente intensidad de color se han clasificado en la misma clase. Por ejemplo, casi no hay distinción entre los árboles y gran parte del cielo.

De acuerdo con los resultados para el caso $n = 2$, el Algoritmo 1 aplica de nuevo fuzzy c-means para obtener una nueva partición con $n = 3$, es decir tres clases. Calculando luego los valores de la relación de proximidad se obtienen los resultados presentados en la Tabla 4.1.

De acuerdo con los valores obtenidos, como $w(\varphi_3^T, \delta_3^T) = 0{,}84 > w(\varphi_n^T, \sigma_n^T) = 0{,}23$, la partición con tres clases es una partición de *calidad.* Por lo tanto, el algoritmo procede a evaluar la relevancia de las clases (Paso 10) considerando todas las particiones de 2 clases C_1,C_2, C_3 y calculando los correspondientes valores para w.

w	ϕ_3^T	σ_3^T	δ_3^T	φ_3^T
ϕ_3^T	1	0,84	0,93	0,17
σ_3^T	0,84	1	0,17	0,23
δ_3^T	0,93	0,17	1	0,84
φ_3^T	0,17	0,23	0,84	1

Tabla 4.1: *Relación de proximidad w*

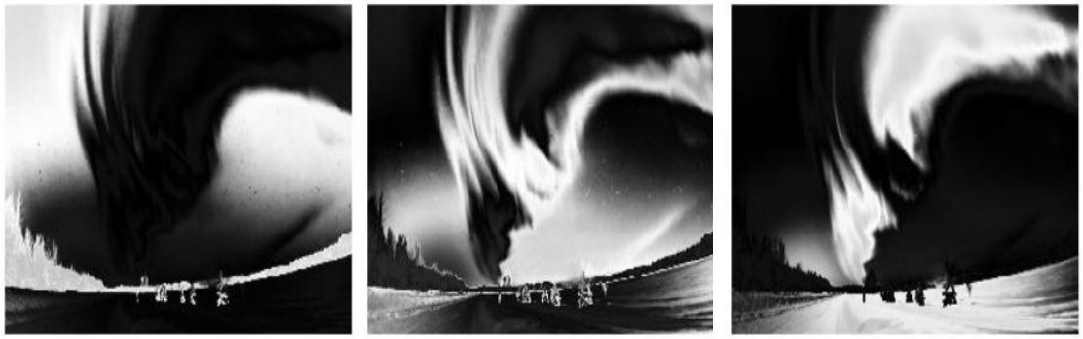


Figura 4.3: *Clases obtenidas por la aplicación del algoritmo fuzzy c-means para $c = 3$ en aurora boreal*
(izquierda: clase 1, centro: clase 2 y derecha: clase 3). La escala de grises representa los grados de pertenencia de cada píxel a cada clases, donde negro= 0 y blanco =1.

Para C_3 se obtiene $k_3 = w(\varphi_2^T, \phi_2^T) = 0{,}42 > w(\varphi_2^T, \sigma_2^T) = 0{,}27 = s_3$. Para C_2 se obtiene $k_2 = w(\varphi_2^T, \phi_2^T) = 0{,}39 > w(\varphi_2^T, \sigma_2^T) = 0{,}36 = s_2$ y para C_1 los resultados obtenidos son, $k_1 = w(\varphi_2^T, \phi_2^T) = 0{,}41 > w(\varphi_2^T, \sigma_2^T) = 0{,}28 = s_1$.

Por lo tanto, aplicando el *criterio de relevancia* se determina que todas las clases son de alta relevancia, es decir, la partición obtenida es de *alta calidad*. La segmentación de la imagen correspondiente a tres clases es presentada en la Figura 4.3.

Considerando la visualización de la segmentación obtenida, por ejemplo, existe una mejor distinción entre los arboles y el cielo en comparación a una partición con 2 clases. Por lo tanto, con una partición borrosa de 3 clases para la imagen seleccionada, se obtiene una buena comprensión en términos de la intensidad del color de los objetos.

En una primera fase, una forma de corroborar los resultados obtenidos sobre la imagen particular, es aplicando de nuevo una segmentación a través de fuzzy c-means para obtener una partición con $n = 4$ clases. Realizado este proceso y calculando

los valores correspondientes a la relación de proximidad se obtuvieron los siguientes resultados, $r = w(\varphi_4^T, \delta_4^T) = 0{,}87 > s = w(\varphi_4^T, \sigma_4^T) = 0{,}24 > k = w(\varphi_4^T, \phi_4^T) = 0{,}12$, con lo cual la partición es de *calidad*. Sin embargo, evaluando la relevancia de sus clases se encuentra que, $k_1 = w(\varphi_3^T, \phi_3^T) = 0{,}3 < w(\varphi_3^T, \sigma_3^T) = s_1 = 0{,}41$, con lo cual, existe una clase que no es relevante.

Adicionalmente, con el fin de ilustrar la robustez de los resultados obtenidos, el Algoritmo 1 ha sido aplicado con diferentes ternas de De Morgan, denominadas Terna 1 y Terna 2, las cuales son presentadas en la Tabla 4.2. En ambos casos se ha trabajado con la negación fuerte $N(x) = 1 - x$.

Rules	$\phi_n(\mu_1(x), \ldots, \mu_n(x))$	$\varphi_n(\mu_1(x), \ldots, \mu_n(x))$	$N(x)$
Terna 1	$\min(\mu_1(x), \ldots, \mu_n(x))$	$\max(\mu_1(x), \ldots, \mu_n(x))$	$1 - x$
Terna 2	$\dfrac{\sqrt{\prod_{i=1}^{n} \mu_i(x)}}{\sqrt{\prod_{i=1}^{n} \mu_i(x) + 1 - \prod_{i=1}^{n} \mu_i(x)}}$	$\dfrac{\sum_{i=1}^{n} \mu_i(x) - \prod_{i=1}^{n} \mu_i(x)}{\sqrt{\prod_{i=1}^{n} (1 - \mu_i(x)) + \sum_{i=1}^{n} \mu_i(x) - \prod_{i=1}^{n} \mu_i(x)}}$	$1 - x$

Tabla 4.2: *Dos ternas De Morgan*

Los valores obtenidos para una partición con tres clases reflejan, independientemente de la terna de De Morgan empleada, la alta relevancia de sus clases. Los resultados son presentados en la Tabla 4.3.

$n = 3$	Terna 1	Terna 2
C	$r = 0{,}84,\ s = 0{,}38$	$m = 0{,}77, s = 0{,}57$
C_1	$r = 0{,}68, s = 0{,}36, k = 0{,}5$	$r = 0{,}69, s = 0{,}51, k = 0{,}65$
C_2	$r = 0{,}63, s = 0{,}33, k = 0{,}44$	$r = 0{,}62, s = 0{,}51, k = 0{,}59$
C_3	$r = 0{,}67, s = 0{,}36, k = 0{,}51$	$r = 0{,}68, s = 0{,}51, k = 0{,}67$

Tabla 4.3: *Indicadores de calidad y relevancia para una partición con tres clases y dos ternas diferentes*

Lo anterior confirma que la mejor segmentación para la imagen tratada, corresponde a una partición borrosa formada por tres clases mediante el algoritmo fuzzy c-means.

4.2. Cálculo del número óptimo de clases

Con el fin de aplicar la propuesta sobre el cálculo del número óptimo de clases bajo un conjunto de particiones borrosas, se ha seleccionado la imagen presentada en la Figura 4.4. De nuevo se considera el problema de clasificación no supervisado de obtener clases de píxeles similares empleando el algoritmo fuzzy c-means. Al igual que en la sección anterior, la imagen seleccionada adquiere un interés particular en el marco de la propuesta planteada debido a que algunos bordes no se encuentran claramente definidos y la separación entre objetos no es nítida.

Figura 4.4: *Imagen extraída de Berkeley Segmentation Dataset (BSDS500)* [77].

Se considera la tripleta recursiva definida por

1. $\phi_n(\mu_1(x),\ldots,\mu_n(x)) = \dfrac{3\prod_{k=1}^{n}\mu_k(x)}{1+2\prod_{k=1}^{n}\mu_k(x)}$
2. $\varphi_n(\mu_1(x),\ldots,\mu_n(x)) = \dfrac{1-\left(\prod_{k=1}^{n}(1-\mu_k(x))\right)}{1+2\prod_{k=1}^{n}(1-\mu_k(x))}$
3. $N(\mu(x)) = 1-\mu(x)$.

Se emplea la función de disimilitud sobre G_T dada por $d(\theta_m^T,\lambda_m^T) = |\theta_m^T - \lambda_m^T|$. Para resolver el problema de optimización planteado en 3.2, es decir,

$$\arg\min_{n}\left\{\int_{-a}^{a}\left|p_c(\lambda)-p_{\bar{d}_n}(\lambda)\right|d\lambda\right\}$$

se han seleccionado los valores de $a = 2$ y $n = 5$. Por lo tanto, se considera el conjunto $\mathbb{C}_4 = \{C_2, ..., C_5\}$.

De acuerdo con la Definición 24, para determinar el número óptimo, se calculan los correspondientes polinomios característicos de las matrices obtenidas por la función d para cada n. Posteriormente, se calcula $\int_{-2}^{2}\left|p_c(\lambda)-p_{\bar{d}_n}(\lambda)\right|d\lambda$ para cada n. En el caso particular de la imagen seleccionada, la Tabla 4.4 presenta los resultados obtenidos.

Partición	Polinomios característicos	$\int_{-2}^{2}\left\|p_c(\lambda)-p_{\bar{d}_n}(\lambda)\right\|$
2 Clases	$x^4 - 1{,}2x^2 - 1{,}05 \cdot 10^{-15}x - 2{,}3 \cdot 10^{-31}$	14.91
3 Clases	$x^4 - 2{,}34x^2 - 1{,}06x - 0{,}119$	8.69
4 clases	$x^4 - 2{,}35x^2 - 1{,}17x - 0{,}145$	8.65
5 Clases	$x^4 - 1{,}34x^2 - 0{,}74x - 0{,}102$	13.82

Tabla 4.4: *Polinomios característicos de* $2, 3, 4$ *y* 5 *clases para la Figura 4.4.*

De igual forma, la Figura 4.5 presenta los polinomios característicos obtenidos para $A_{\bar{d}_n}$ con $n = 2, 3, 4, 5$ y el polinomio característico del umbral, es decir, $p(\lambda) = x^4 - 4x^2$.

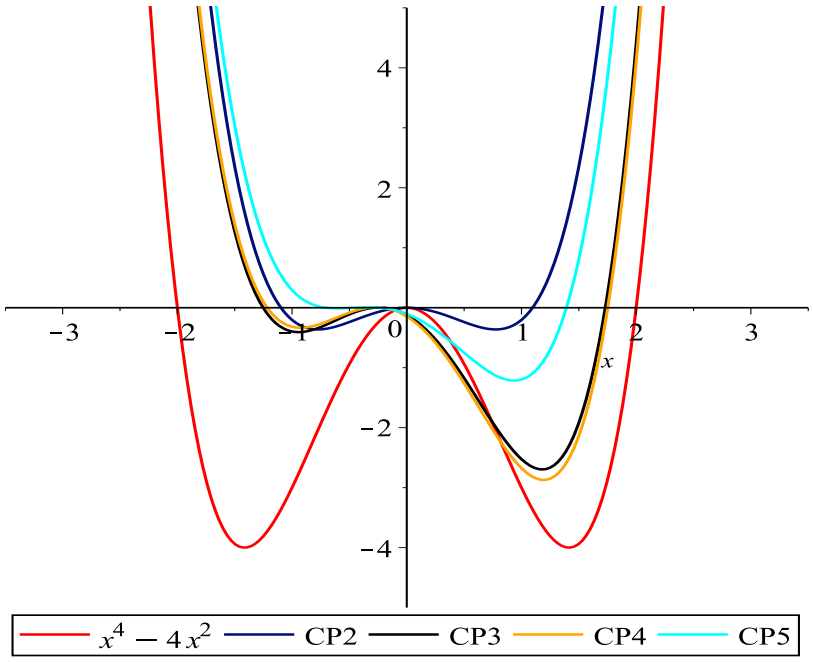


Figura 4.5: *Polinomios característicos para el umbral y para las matrices* $D_{\bar{d}_2}$, $D_{\bar{d}_3}$, $D_{\bar{d}_4}$ *y* $D_{\bar{d}_5}$.

Se han denotado por CP2, CP3, CP4, y CP5 respectivamente

Según el procedimiento anterior, se establece que $n = 4$, es decir, el conjunto con cuatro clases obtenidas a través del algoritmo fuzzy c-means, es el número óptimo de clases para la Figura 4.4. En otras palabras, la partición borrosa con cuatro clases es la que mejor clasifica píxeles similares.

Las clases correspondientes de dicha partición borrosa para la Figura 4.4 se presentan en la Figura 4.6, 4.7, 4.8 y 4.9. La escala de grises representa el grado de pertenencia de cada píxel a la clase, donde negro = 0 y blanco = 1. A partir de la visualización de las clases se interpreta la segmentación obtenida como se describe a continuación:

- Las clases presentadas en las Figuras 4.6 - 4.7 dividen la imagen original en dos regiones. La clase obtenida en la Figura 4.6 se ha denominado como "clase fondo negro" y la clase obtenida en la Figura 4.7 se ha denominado "clase fondo arena".

Por lo tanto, se concluye que en general, la Figura 4.4 se encuentra compuesta por una parte completamente negra y otra parte con diferentes intensidades de color, correspondiente a lo que se ha denominado “fondo arena”.

- Sin embargo, la “clase fondo arena” se ha segmentado en dos nuevas clases (que no contienen la “clase fondo negro”): la clase obtenida en la Figura 4.8 se ha denominado “clase objeto” y la clase obtenida en la Figura 4.9 se ha denominado “clase arena sin objeto”.

- Estas dos últimas clases, naturalmente, comparten intensidades de color similares, lo cual se ha hace evidente en la “clase fondo arena”. Sin embargo, aunque hay clases que comparten píxeles de intensidad similar, la segmentación de la “clase fondo arena” en dos clases adicionales se interpreta como la presencia de intensidades de color únicas del objeto y las intensidades de color únicas de la arena.

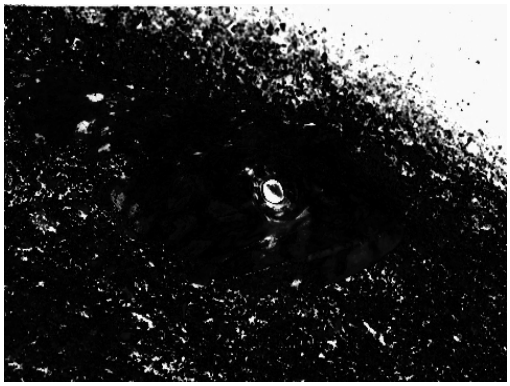


Figura 4.6: *Clase fondo negro. Parte de color negro y las restantes intensidades de color similares*

Figura 4.7: *Clase fondo arena. Parte representando la arena y las intensidades de color similares.*

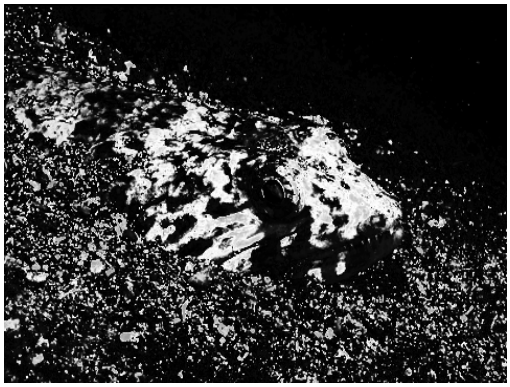


Figura 4.8: *Clase objeto. Objeto y sus propias intensidades de color.*

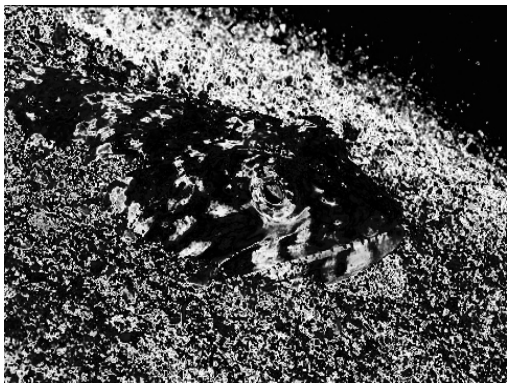


Figura 4.9: *Clase arena sin objeto.*

Finalmente, para evidenciar las consistencia del índice propuesto se han seleccionado y aplicado sobre la Figura 4.4 el FCS con negación de Sugeno-Yager compuestos por las siguientes reglas recursivas:

Sistema de clasificación borrosa (FCS) conformado por:

1. $\phi_n(\mu_1(x), \ldots, \mu_n(x)) = \prod_{i=1}^{n} \mu_i(x)$
2. $\varphi_n(\mu_1(x), \ldots, \mu_n(x)) = \dfrac{\prod_{i=1}^{n}\left(1+2(\mu_i(x))\right) - \prod_{i=1}^{n}(1-\mu_i(x))}{\prod_{i=1}^{n}\left(1+2(\mu_i(x))\right) + \prod_{i=1}^{n}(1-\mu_i(x))}$
3. $N(\mu(x)) = \frac{1-x}{1+2x}$

Obteniendo los resultados presentados en la Tabla 4.5, los cuales coinciden en la determinación del número óptimo de la Tabla 4.4.

Clases	FCS $\int_{-2}^{2} \lvert p_c(\lambda) - p_{\tilde{d}_n}(\lambda) \rvert$
2 clases	10,879
3 clases	12,396
4 clases	10,878
5 clases	10,991
6 clases	11,029

Tabla 4.5: *Resultados de aplicar FCS con negación de Sugeno-Yager en el cálculo de número óptimo de clases sobre la Figura 4.4 para* $2, 3, 4, 5$ *y* 6 *clases.*

4.3. Comparación de índices

En las secciones 4.1 y 4.2 se han presentado con detalle la aplicación de los criterios y definiciones planteadas tanto para la evaluación de una partición borrosa como para la determinación del número óptimo de clases en un proceso de clasificación borrosa. En la presente sección se busca comparar los resultados de la propuesta tanto con un conjunto de índices ya existentes y que son ampliamente usados en la literatura, como con una clasificación proporcionada por un experto precisamente en el análisis de imágenes. En adelante los índices empleados se denominarán *índices clásicos*.

Con el fin de comparar los resultados obtenidos, el trabajo se ha divido en dos partes. La primera parte corresponde a la aplicación de los criterios de calidad y relevancia en la Figura 4.4 y al cálculo del número óptimo de clases en la Figura 4.1. En ambos casos se comparan los resultados con el conjunto de índices clásicos. Lo anterior se realiza con el fin de complementar el análisis de las figuras en mención con todos los elementos de la propuesta.

Para la segunda parte, se ha seleccionado una tomografía axial computarizada y sobre ella se aplican: los criterios de calidad y relevancia, el número óptimo de clases, el conjunto de índices clásicos, y adicionalmente, se emplea una segmentación realizada previamente por expertos y por un software especializado de análisis de imágenes médicas.

4.3.1. Comparación sobre imágenes a color

En primera instancia, con el fin de complementar el análisis sobre la imagen de la *aurora boreal*, es necesario establecer el número óptimo de clases empleando los polinomios característicos. Por lo tanto, se ha seleccionado el mismo operador aplicado sobre la imagen (Figura 4.1) para el desarrollo del Algoritmo 1. La Tabla 4.6 presenta los resultados de calcular el número óptimo de clases sobre la Figura 4.1.

Partición	Polinomios característicos	$\int_{-2}^{2} \lvert p_c(\lambda) - p_{\bar{d}_n}(\lambda) \rvert$
2 Clases	$x^4 - 2{,}23x^2 - 0{,}47x + 0{,}35$	10.85
3 Clases	$x^4 - 1{,}99x^2 - 0{,}41x + 0{,}36$	12.16
4 Clases	$x^4 - 2{,}27x^2 - 0{,}37x + 0{,}51$	11.28
5 Clases	$x^4 - 2{,}24x^2 - 0{,}38x - 0{,}5$	11.35

Tabla 4.6: *Polinomios característicos para* $2, 3, 4$ *y* 5 *clases para la Figura 4.1*

De acuerdo con los valores obtenidos se establece que el número óptimo de clases para la imagen 4.1, según el índice propuesto, corresponde con una partición de 2 clases.

De igual forma, se realizó el cálculo del conjunto de índices presentados en la sección 1.4.1.2 correspondientes al coeficiente de partición (CP), el índice de entropía (IE), el coeficiente de partición modificado (CPM), el índice de Xie-Beni (XB) y, el índice de Kwon (Kwon), cuyos resultados se presentan en la Tabla 4.7.

Índice	2 clases	3 clases	4 clases	5 clases	Criterio
CP	0,8	0,71	0,618	0,57	Máximo valor
CPM	0,615	0,56	0,49	0,471	Máximo valor
IE	0,31	0,52	0,71	0,82	Mínimo Valor
XB	0,09	0,11	0,14	0,147	Mínimo Valor
Know	5599,1	6490,94	8516,8	8622,26	Mínimo Valor

Tabla 4.7: *Índices clásicos de validación de particiones borrosas y criterios para la segmentación de la Figura 4.1 empleando fuzzy c-means con 2, 3, 4 y 5 clases.*

De acuerdo con los valores encontrados para cada índice de la Tabla 4.7, se determina que una partición con 2 clases es la que mejor comportamiento presenta frente a los criterios de cada índice.

En resumen, la Tabla 4.8 presenta los resultados obtenidos en el análisis de la Figura 4.1, de acuerdo con las mediciones realizadas.

Método de evaluación	Resultado
Criterios de calidad y relevancia	Partición con 3 clases
Número óptimo de clases (Polinomios característicos)	Partición con 2 clases
Índices clásicos (CP), (CPM), (IE), (XB), (Kwon)	Partición con 2 clases

Tabla 4.8: *Resultados de la evaluación de particiones borrosas sobre la Figura 4.1 empleando fuzzy c-means con 2, 3, 4 y 5 clases.*

A partir de lo anterior, se observa una diferencia de resultados entre el conjunto de índices (número óptimo e índices clásicos) y los criterios de calidad y relevancia propuestos. En este punto es importante resaltar que de acuerdo con lo expuesto en la Sección 4.1, una partición con 4 clases para la imagen analizada bajo los criterios de calidad y relevancia no es considerada una partición de *alta calidad*.

A continuación, se realiza el análisis de la Figura 4.4 con el fin de establecer algunas conclusiones sobre las diferencias y/o similitudes encontradas, teniendo en cuenta todas las mediciones realizadas para ambas imágenes.

En la Sección 4.2, se estableció el número óptimo de clases para la Figura 4.4. Por lo tanto, con el fin de continuar con el proceso se requiere implementar el Algoritmo 1 determinando la calidad y relevancia de las clases. En este sentido, se han seleccionado los siguientes parámetros como entradas del algoritmo.

- Sistema de clasificación borrosa (FCS) conformado por:

  1. $\phi_n(\mu_1(x), \ldots, \mu_n(x)) = \prod_{i=1}^{n} \mu_i(x)$
  2. $\varphi_n(\mu_1(x), \ldots, \mu_n(x)) = \dfrac{\prod_{i=1}^{n}\left(1+2(\mu_i(x))\right) - \prod_{i=1}^{n}(1-\mu_i(x))}{\prod_{i=1}^{n}\left(1+2(\mu_i(x))\right) + \prod_{i=1}^{n}(1-\mu_i(x))}$
  3. $N(\mu(x)) = \frac{1-x}{1+2x}$

- Algoritmo de clasificación iterativo no jerárquico: Fuzzy c-means
- Función de proximidad: $w(\theta_m^T, \lambda_m^T) = 1 - |\theta_m^T - \lambda_m^T|$.
- Número máximo de clases a considerar: $z = 8$

Una vez implementado, el Algoritmo 1 encontró una partición de *alta calidad*. La Tabla 4.9 presenta los resultados obtenidos.

Partición	Resultados Algoritmo 1	Criterio
2 clases	$s = 0{,}1 < k = 0{,}17$	Partición de baja calidad
3 clases	$k = 0{,}19 < s = 0{,}26 < r = 0{,}82$	Partición de calidad
Relevancia clase 1	$k = 0{,}38 > s = 0{,}25$	Clase relevante
Relevancia clase 2	$k = 0{,}34 > s = 0{,}2$	Clase relevante
Relevancia clase 3	$k = 0{,}47 > s = 0{,}33$	Clase relevante
		Partición de alta calidad encontrada

Tabla 4.9: *Resultados obtenidos por la aplicación del Algoritmo 1 para la Figura 4.4.*

Lo anterior indica que bajo los criterios de calidad y relevancia, la partición con tres clases corresponde a una partición de *alta calidad*.

Continuando con el análisis, se realiza el cálculo de los índices clásicos sobre la Figura 4.4. La Tabla 4.10 presenta los valores para los índices clásicos empleados.

Índice	2 clases	3 clases	4 clases	5 clases	Criterio
CP	0,80	0,75	0,71	0,68	Máximo valor
CPM	0,601	0,628	0,622	0,6007	Máximo valor
IE	0,32	0,44	0,54	0,63	Mínimo Valor
XB	0,106	0,093	0,103	0,113	Mínimo Valor
Know	16382,05	14398,3	15924,67	17,627,74	Mínimo Valor

Tabla 4.10: *Índices clásicos de validación de particiones borrosas y criterios para la segmentación de la Figura 4.4 empleando fuzzy c-means con 2, 3, 4 y 5 clases.*

Los resultados de la Tabla 4.10 muestran que no existe uniformidad en la selección de la mejor partición por parte de los índices clásicos. De igual forma, al comparar los resultados de los criterios de calidad y relevancia con el número óptimo de clases, se encuentran diferencias frente al resultado final de la evaluación. La Tabla 4.11 presenta el resumen del análisis sobre la Figura 4.4, de acuerdo con las mediciones realizadas.

Método de evaluación	Resultado
Criterios de calidad y relevancia	Partición con 3 clases
Número óptimo de clases (Polinomios característicos)	Partición con 4 clases
Índices clásicos (CP), (IE),	Partición con 2 clases
Índices clásicos (CPM), (XB), (Kwon)	Partición con 3 clases

Tabla 4.11: *Resultados de la evaluación de particiones borrosas sobre la Figura 4.4 empleando fuzzy c-means con 2, 3, 4 y 5 clases.*

Debido a la amplia variedad de resultados encontrados y, revisando en detalle algunos valores obtenidos, se ha decidido aplicar el Algoritmo 1 adicionando una clase a la partición de *alta calidad* encontrada, es decir, aplicar los criterios de calidad y relevancia a una partición con 4 clases. Lo anterior, se justifica particularmente en lo siguiente: 1) una comparación entre los valores obtenidos con el coeficiente de partición modificado (CPM) y con el número óptimo de clases (polinomios característicos), permiten establecer que los valores hallados para particiones con 3 y 4 clases se encuentran bastante próximos, al menos, más cerca que otros índices para distintas particiones; 2) determinar o descartar que exista una partición adicional que cumpla con las condiciones de *alta calidad*.

Después de aplicar el Algoritmo 1 específicamente a una partición con 4 clases sobre la Figura 4.4, se obtienen los resultados presentados en la Tabla 4.12.

Partición	Resultados Algoritmo 1	Criterio
4 clases	$k = 0{,}09 < s = 0{,}12 < r = 0{,}9$	Partición de calidad
Relevancia clase 1	$s = 0{,}28 < k = 0{,}33$	Clase relevante
Relevancia clase 2	$k = 0{,}35 > s = 0{,}34$	Clase relevante
Relevancia clase 3	$k = 0{,}24 > s = 0{,}18$	Clase relevante
Relevancia clase 4	$k = 0{,}241 > s = 0{,}22$	Clase relevante
		Partición de alta calidad encontrada

Tabla 4.12: *Resultado de la evaluación de la calidad y relevancia para una partición de 4 clases sobre la Figura 4.4 a través del Algoritmo 1.*

De acuerdo con lo anterior, se establece que una partición con 4 clases es igualmente una partición de *alta calidad*. Por lo tanto, de acuerdo con el *criterio de com-*

paración presentado en la Sección 3.2, el valor de r determina entre dos particiones de *alta calidad*, cuál es la mejor. Para el caso particular, se tiene que la partición con 3 clases tiene un $r = 0{,}82$ y la partición con 4 clases presenta un $r = 0{,}9$. Esto se interpreta como que la partición con 4 clases tienen más altos grados de cubrimiento sin considerar la redundancia de las clases, con lo cual, la partición con 4 clases es mejor.

El análisis desarrollado sobre las dos imágenes tratadas en la presente sección, permite establecer, en principio, las siguientes inferencias: 1) Los criterios de calidad y relevancia aportan información adicional a la evaluación de una partición borrosa, en contraste con los índices clásicos y el índice a partir de polinomios característicos. Específicamente, permite una revisión del comportamiento de las clases que conforman una partición. En principio debido a esto, es posible explicar la diferencia con los resultados proporcionados por otros índices, como se observa en el análisis de la Figura 4.1; 2) bajo un primer análisis, el número óptimo de clases a partir de los polinomios característicos presenta un comportamiento similar a los índices clásicos tratados, en parte debido al uso de un umbral nítido. Sin embargo, dado que las información contenida en el polinomio característico y, por ende, la información correspondiente a la matriz de disimilitud relaciona varias propiedades de la partición de forma simultanea (redundancia, agrupamiento, no redundancia y no solapamiento), es posible que este índice permita una visión más amplia de las particiones, como puede estar ocurriendo en la Figura 4.4.

Con el fin de ir reforzando y validando las inferencias establecidas, a continuación se realiza nuevamente un análisis con todas la mediciones empleadas sobre una imagen médica. Para dicha imagen, se cuenta con un proceso de validación previa por parte de un experto.

4.3.2. Comparación sobre imagen tomográfica

Con el fin de continuar con el proceso de validación y comparación de la propuesta planteada, se adiciona un elemento que permite tener una visión más amplia de los resultados obtenidos: la validación a partir de expertos. De acuerdo con esto, se ha seleccionado una de las veinte imágenes de la base de datos proporcionada en [58] y cuya finalidad ha sido la segmentación pulmonar para la identificación de infecciones a partir de tomografías computarizadas de COVID-19. Los resultados del estudios sobre dicha base de datos son publicados en [57], y para su desarrollo, los autores de tal estudio han contado con la validación por parte de tres radiólogos expertos para la posterior segmentación a partir de técnicas de *deep learning*.

Los exámenes tomográficos generan una amplia cantidad de imágenes denominadas cortes o secciones, sobre las cuales se realiza el diagnóstico final. La Figura 4.10 presenta cuatro de las secciones de un mismo examen.

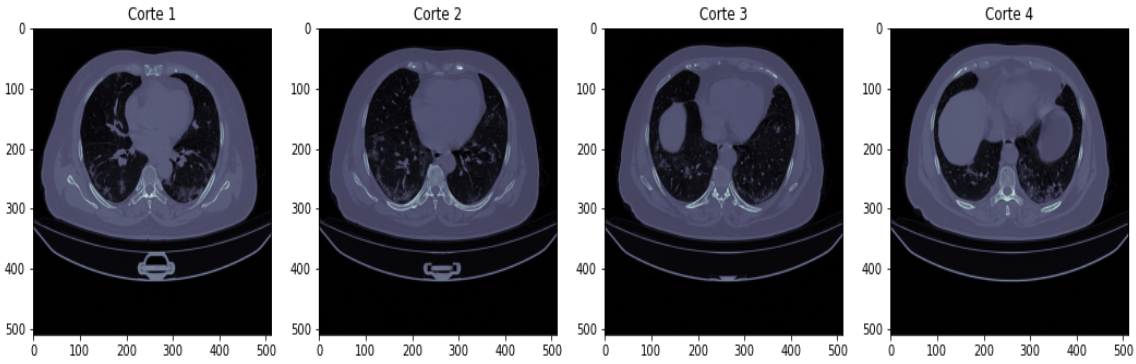


Figura 4.10: *Cortes de examen tomográfico. Visualización a partir de paquete NiBabel para Python.*

Imagen extraída de [58].

En el caso particular de la presente tesis, se empleará únicamente el corte 4 de dicha imagen, la cual se presenta el Figura 4.11.

Figura 4.11: *Tomografía computarizada pulmón COVID-19. Visualización a partir del software Fiji(ImajeJ).*

Imagen extraída de [58].

En el estudio desarrollado en [57], el primer paso es la identificación de los pulmones y de la región correspondiente a la infección por parte de los expertos. Posteriormente, se proporcionan mascaras que permiten diferenciar los pulmones, la región afectada y la región correspondiente a otros órganos. [1] La Figura 4.12 presenta la segmentación obtenida sobre la Figura 4.11 a partir de las mascaras proporcionadas por los expertos.

[1]Código obtenido de https://www.kaggle.com/andrewmvd/covid-19-ct-scans-getting-started-COVID-Getting-Started.

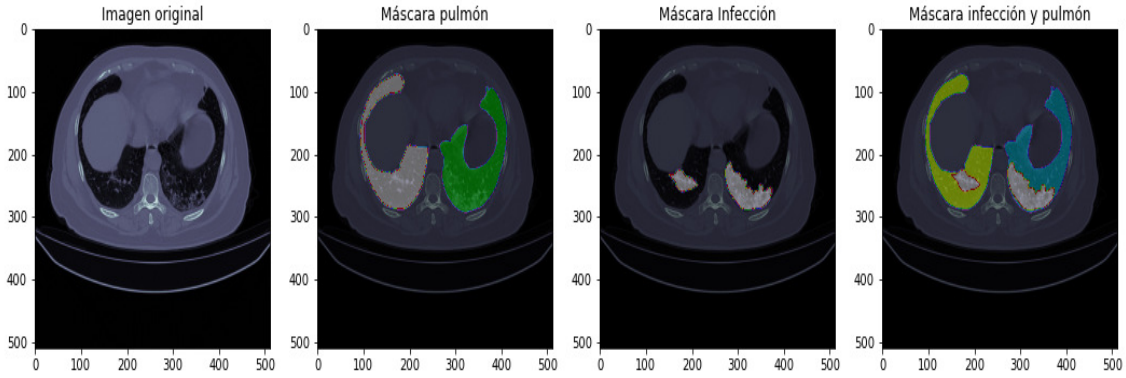


Figura 4.12: *Segmentación de tomografía.*

Sin embargo, a partir de la imagen original, un elemento relevante en el desarrollo de la presente tesis es la identificación de diferentes densidades en la región correspondiente a *otros órganos*. Dicha región comprende elementos que no son nítidos y su identificación implica bordes no definidos claramente. Este hecho se valida a partir de un procesamiento de la imagen con la ayuda del software especializado *Fiji(ImageJ)*, el cual procesa y analiza imágenes propias del campo de la medicina. Dicho software realiza una segmentación de las tomografías según sus densidades tomográficas, lo cual permite identificar elementos de diferentes naturaleza según su composición (agua, calcio, aire, grasas, entre otras). La Figura 4.13 presenta los resultados de la segmentación, en la cual se hacen visibles, en la zona de *otros órganos*, elementos de diferente naturaleza y que permiten determinar por lo menos una clase o región adicional en el interior de dicha zona.

Figura 4.13: *Segmentación de tomografía a través de software Fiji(ImageJ) reconociendo densidades tomográficas.*

Con base en lo anterior, se emplean las regiones validadas por el estudio desarrollado en [57] y el procesamiento a partir del software especializado, como *validador experto* de la segmentación obtenida a partir de un proceso de clasificación borrosa. Específicamente, se establece que una partición borrosa de calidad debe permitir, a través de la visualización de sus clases, identificar las siguientes zonas: el fondo negro de la imagen, los pulmones, las zonas infectadas, y las diferencias de densidad en la zona de *otros órganos*.

Por lo tanto, se busca validar los resultados de la propuesta presentada en la tesis correspondiente al Algoritmo 1 y la determinación del número óptimo de clases, en la segmentación borrosa de la imagen mostrada en la Figura 4.11 empleando fuzzy c-means. Los resultados obtenidos se contrastan con lo que se ha denominado *validador experto* y a su vez, se comparan con el conjunto de índices clásicos.

De acuerdo con lo anterior, para la aplicación del Algoritmo 1 sobre la Figura 4.11, se consideran las siguientes entradas:

- Sistema de clasificación borrosa (FCS) conformado por:

  1. $\phi_n(\mu_1(x), \ldots, \mu_n(x)) = \prod_{i=1}^{n} \mu_i(x)$
  2. $\varphi_n(\mu_1(x), \ldots, \mu_n(x)) = \frac{\prod_{i=1}^{n}\left(1+2(\mu_i(x))\right) - \prod_{i=1}^{n}(1-\mu_i(x))}{\prod_{i=1}^{n}\left(1+2(\mu_i(x))\right) + \prod_{i=1}^{n}(1-\mu_i(x))}$
  3. $N(\mu(x)) = \frac{1-x}{1+2x}$

- Algoritmo de clasificación iterativo no jerárquico: Fuzzy c-means

- Función de proximidad: $w(\theta_m^T, \lambda_m^T) = 1 - |\theta_m^T - \lambda_m^T|$.

- Número máximo de clases a considerar: $z = 8$.

Una vez aplicado el Algoritmo 1 se ha encontrado una partición borrosa de *alta calidad* la cual está compuesta por tres clases, es decir, dicha partición es de calidad y todas sus clases son relevantes. Los valores obtenidos de acuerdo con el Algoritmo 1 se presentan el Tabla 4.13.

w	ϕ_3^T	σ_3^T	δ_3^T	φ_3^T
ϕ_3^T	1,000	0,927	0,969	0,076
σ_3^T	0,927	1,000	0,076	0,107
δ_3^T	0,969	0,076	1,000	0,927
φ_3^T	0,076	0,107	0,927	1,000

Tabla 4.13: *Relación de proximidad w para la partición de tres clases sobre la Figura 4.11 a través del Algoritmo 1.*

En particular, el orden determinado por los valores $r = 0{,}927 > s = 0{,}107 > k = 0{,}076$, determinan que la partición borrosa con tres clases, es una partición de *calidad*. A continuación, el Algoritmo 1 estableció los valores de relevancia para las tres clases, con lo cual el algoritmo ha finalizado. Dichos valores son presentados en la la Tabla 4.14.

Clases	Valores para medir relevancia
Clase 1	$s = 0,17$, $k = 0{,}68$
Clase 2	$s = 0{,}09$, $k = 0{,}23$
Clase 3	$s = 0{,}09$, $k = 0{,}18$

Tabla 4.14: *Medición de relevancia de las clases*

A continuación se procede con la visualización de las clases para contrastar con el *validador experto*. La Figura 4.14 presenta los resultados de la segmentación de la imagen original para la partición de *alta calidad*.

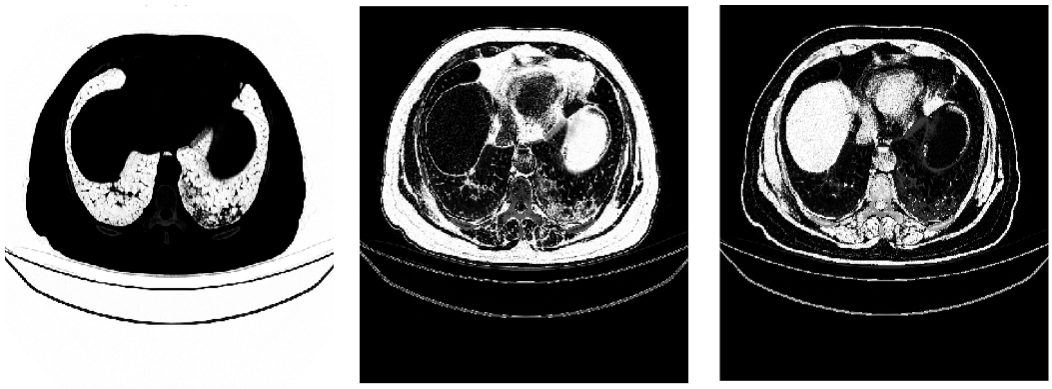


Figura 4.14: *Clases obtenidas por la aplicación del algoritmo fuzzy c-means para* $c = 3$ *en tomografia*

(izquierda: clase 1, centro: clase 2 y derecha: clase 3). La escala de grises representa los grados de pertenencia de cada píxel a cada clases, donde negro = 0 y blanco = 1.

En la Figura 4.14, se observa que la *clase 1* contiene el fondo de la imagen original junto con los píxeles de igual intensidad correspondientes a los pulmones. La *clase 2* corresponde a ciertas regiones de la zona denominada *otros órganos*, donde se puede ver que dicha clase contiene las zonas infectadas de los pulmones, identificándolas con intensidades similares de otros órganos. Finalmente, la *clase 3* contiene las regiones de otros órganos con intensidades diferentes a los pulmones y las regiones infectadas. Las clases 2 y 3 permiten comprobar la premisa inicial sobre la existencia de regiones poco nítidas y, por lo tanto, de difícil clasificación.

A continuación, se contrastan los resultados obtenidos con el conjunto de índi-

ces clásicos: el Coeficiente de Partición (CP), el Coeficiente de Partición Modificado (CPM), el Índice de Entropía (IE), el índice de Xie Beni(XB) y el Índice Know. La Tabla 4.15 presenta los resultados de los índices para particiones con dos, tres y cuatro clases.

Índice	2 clases	3 clases	4 clases	Criterio
CP	0,97	0,92	0,87	Máximo valor
CPM	0,94	0,88	0,83	Máximo valor
IE	0,05	0,15	0,22	Mínimo Valor
XB	0,01	0,086	0,165	Mínimo Valor
Know	4187,1	22796,81	43337,75	Mínimo Valor

Tabla 4.15: *Índices de validación de particiones borrosas y criterios para la segmentación de la Figura 4.11 empleando fuzzy c-means con 2, 3 y 4 clases.*

Con base en los índices calculados y los criterios correspondientes para cada uno de ellos, se identifica que todos los índices determinan la partición con 2 clases como la partición de más alta calidad. La Figura 4.15 presenta la partición con dos clases.

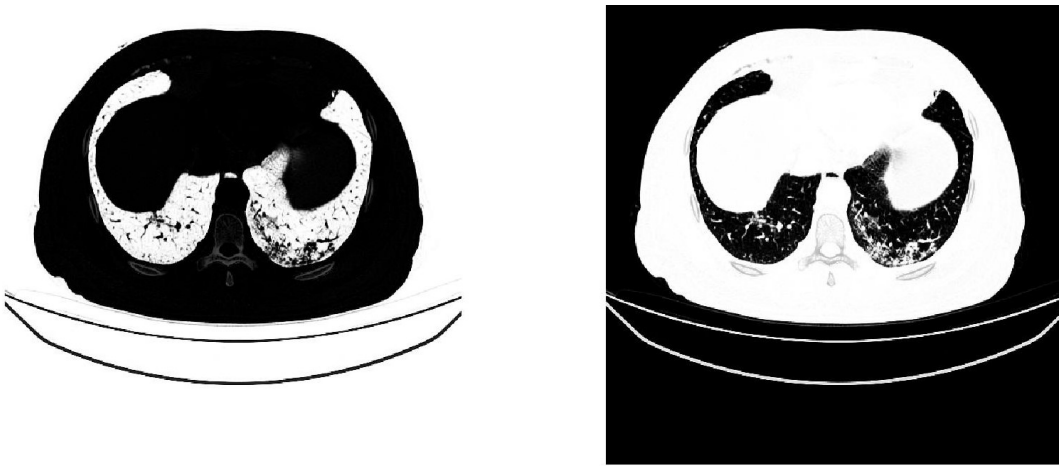


Figura 4.15: *Clases obtenidas por la aplicación del algoritmo fuzzy c-means para* $c = 2$ *en tomografía*

(izquierda: clase 1, derecha: clase 2). La escala de grises representa los grados de pertenencia de cada píxel a cada clases, donde negro= 0 y blanco =1.

A partir de la Figura 4.15, se identifica una amplia perdida de información con una

partición borrosa de dos clases debido a que, en la zona de *otros órganos*, no se están reconociendo los elementos con diferentes densidades. La clase 2 corresponde a dicha zona y no se distinguen valores de pertenencia intermedios que permitan reconocer los elementos allí existentes. Únicamente aparecen diferenciadas la zona de los pulmones junto con el fondo y la zona infectada. Lo anterior soporta las ventajas de la propuesta presentada en esta tesis con relación a los criterios de calidad y relevancia para incluir información del contexto que de otra forma sería ignorada.

Al igual que en la sección anterior, las diferencias de los resultados encontrados en contraste con los índices clásicos, están evidenciando que, la evaluación tanto de la partición en conjunto como cada una de sus clases por separado a partir de los criterios de calidad y relevancia, están proporcionando información que difícilmente puede ser evaluada con un solo índice. Específicamente, se están considerando dos propiedades globales de la partición como lo son el solapamiento y el agrupamiento y una propiedad que se considera local, en el sentido de que es evaluada para cada clase. Por lo tanto, no es un desacierto la falta de coincidencia entre los índices y la propuesta sino una ratificación sobre la complejidad de la evaluación de una partición borrosa y su innegable necesidad de emplear distintos mecanismos para determinar su precisión y, sobre todo las necesidades de quien establece el proceso de clasificación.

A continuación, se busca determinar el número óptimo de clases, es decir, del cálculo del polinomio característico de la matriz de disimilitud. Para tal fin, se ha seleccionado el mismo FCS empleado en la aplicación del Algoritmo 1. De igual forma, se aplicará la función de disimilitud sobre G_T dada por $d(\theta_m^T, \lambda_m^T) = |\theta_m^T - \lambda_m^T|$, y se selecciona un máximo de clases de 6. Los resultados se presentan en la Tabla 4.16.

Clases	Polinomios característicos	$\int_{-2}^{2} \lvert p_c(\lambda) - p_{\bar{d}_n}(\lambda) \rvert$
2 clases	$x^4 - 2{,}58x^2 - 5{,}76x + 0{,}73$	10,53
3 clases	$x^4 - 2{,}32x^2 - 0{,}29x + 0{,}67$	11,09
4 clases	$x^4 - 2{,}09x^2 - 0{,}70x + 0{,}39$	11,73
5 clases	$x^4 - 2{,}11x^2 - 0{,}39x + 0{,}40$	11,68
6 clases	$x^4 - 1{,}95x^2 - 0{,}79x - 0{,}03$	11,94

Tabla 4.16: *Polinomios característicos para la matriz de disimilitud sobre la Figura 4.11 para* $2, 3, 4, 5$ *y* 6 *clases.*

De acuerdo con la Tabla 4.16, el número óptimo de clases corresponde, al igual que con lo índices de la Tabla 4.15, a una partición con 2 clases. El índice propuesto aunque genere una medición de la partición en conjunto, sin considerar las clases por separado, está presentando resultados bastante favorables en términos de la evaluación. Lo anterior debido a que, de forma natural y verificando en cierta medida la consistencia del proceso, existen coincidencias con la evaluación dada por otros índices. Por ejemplo, se ha comportado de forma similar cuando todos los índices clásicos seleccionan una misma partición como en el caso de la Figura 4.1 y la tomografía computarizada de Figura 4.11. De forma clara, tanto los criterios de calidad y relevancia como el número óptimo de clases, requieren aún ser aplicados tanto en imágenes de distinta naturaleza y fuente como en otros contextos. Por ejemplo, en la segmentación de imágenes satélites, de imágenes médicas a color o, en su funcionamiento en procesos de clasificación del marketing o en bases de datos.

Trabajos Publicados. Contribuciones de la tesis

Como resultado del proceso investigativo, diversos documentos fueron sometidos a publicaciones con dobles jurados obteniendo su aceptación. Dichos documentos fueron:

- Documento en revistas de impacto indexadas en JCR.

  1. F. Castiblanco, C Franco, J. T. Rodríguez, and J. Montero, *Evaluation of the quality and relevance of a fuzzy partition*. Journal of Intelligent Fuzzy Systems,(Pre-press):1–16, 2020. DOI:10.3233/JIFS-200286. [24]

  2. F. Castiblanco, C Franco, J. T. Rodríguez, and J. Montero, *Are reciprocal fuzzy preferences weak or strict preferences?*. Fuzzy sets and Systems. (en segunda revisión)

- Documentos en revistas de impacto indexadas en SJR.

  1. F. Castiblanco, J. Montero, J. T. Rodríguez, and D. Gómez. *Quality assessment of fuzzy classification: An application to solvency analysis*. Fuzzy Economic Review, 22(1):19–31, 2017. DOI:10.25102/fer.2017.01.02. [28]

- Capítulos o secciones de libro con proceso de doble referee

  1. F. Castiblanco, C. Franco, J. T. Rodríguez. and J. Montero. *Degree of global covering and global overlapping in solvency fuzzy classification*. In León Castro, E. Gil Lafuente A.M. Merigó J.M. Kacprzyk J Blanco Mesa

F., editor, Intelligent and Complex Systems in Economics and Business, volume Pre Press. Springer Nature Switzerland AG, 2020. DOI:10.1007/978-3-030-59191-5. [23]

2. F. Castiblanco, C. Franco, J. Montero, and J. T. Rodríguez. *Optimal number of classes in fuzzy partitions*. In B. Bede, M. Cock, M. Ceberio, Wu,Y. and V. Kreinovich, editors, Fuzzy Techniques: Theory and Applications. NAFIPS 2020. Advances in Intelligent Systems and Computing, volume Pre Press. Springer Nature Switzerland AG, 2020. [25]

3. F. Castiblanco, C. Franco, J. Montero, and J. Rodríguez. Aggregation operatorsto evaluate the relevance of classes in a fuzzy partition. In Kearfott R., BatyrshinI., Reformat M., Ceberio M., and Kreinovich V., editors,Fuzzy Techniques:112 Theory and Applications. IFSA/NAFIPS 2019. Advances in Intelligent Systemsand Computing, volume 1000. Springer Nature Switzerland AG, 2019.
DOI:https://doi.org/10.1007/978-3-030-21920-8_2. [26]

4. F. Castiblanco, C. Franco, J. Montero, and J. T. Rodríguez. Relevance of classes in a fuzzy partition. a study from a group of aggregation operators. In Coelho R. Barreto G., editor, Fuzzy Information Processing. NAFIPS 2018. Communications in Computer and Information Science, volume 831. SpringerNature Switzerland AG, 2018. https://doi.org/10.1007/978-3-319-95312-0_9. [27]

5. F. Castiblanco, D. Gómez, J. Montero, and J. T. Rodríguez. Aggregation toolsfor the evaluation of classifications. InIFSA-SCIS 2017 - Joint 17th World Con-gress of International Fuzzy Systems Association and 9th In-

ternational Confe-rence on Soft Computing and Intelligent Systems, pages 1–5, Otsu, Japan, 2017.IEEE. DOI:10.1109/IFSA-SCIS.2017.8023242. [29]

Conclusiones

A partir de la investigación desarrollada se han alcanzado tres grandes resultados expuestos a continuación:

1. Se estableció una estructura de grupo conmutativo que subyace de los sistemas de clasificación borrosa. A partir de la conformación de dos operadores σ_n^T y δ_n^T que surgen de forma natural para un sistema de clasificación borrosa $(C, \phi_n^T, \varphi_n^T, N)$, el grupo denotado G^T, se extiende tanto para operadores aplicados a los grados de pertenencia de cada elemento de la partición borrosa, como a los grados globales de una partición. Con el fin de garantizar la extensión del grupo se hace necesario que el operador global sea estable por negaciones fuertes.

2. Se ha establecido un método que determina la calidad de una partición borrosa a partir de relaciones de similitud aplicadas a los elementos del grupo conmutativo. Dicho método está constituido por la definición de dos criterios; uno sobre la calidad de una partición y el otro sobre la relevancia de las clases de dicha partición. De igual forma, el método se consolida a través de un algoritmo cuya salida determina lo que se ha definido como una *partición de alta calidad*.

3. Se ha construido un índice que mide en conjunto propiedades de una partición borrosa como el agrupamiento y el solapamiento, en relación a una partición nítida. A partir de tal índice, se ha establecido lo que correspondería a un número óptimo de clases para una partición. Dicho índice se calcula encontrando la solución a un problema de optimización que considera funciones de disimilitud sobre los elementos del grupo conmutativo y una matriz de referencia. Esta propuesta establece, considerando los polinomios característicos de las matrices

de disimilitud, la partición borrosa más cercana a una partición crisp, en tanto lo anterior es señal de mayor agrupamiento y menor solapamiento.

Adicionalmente, los alcances descritos ha sido el fruto de un conjunto de resultados más específicos, pero no menos relevantes. Algunos de ellos se exponen a continuación. En la segunda sección se establecieron las condiciones necesarias bajo las cuales una relación de proximidad y una relación de similitud mantienen invariantes algunos grados. Se definió el concepto de agregación global de una propiedad evaluada sobre una partición borrosa. Se demostró que dado un grupo G es posible establecer una relación de similitud invariante por translaciones, de tal forma que dicha relación conforma un subgrupo borroso normal de $G \times G$. Finalmente, en la misma sección se determinaron las particularidades de los polinomios característicos de matrices de disimilitud sobre los elementos del grupo conmutativo.

En la tercera sección, se estableció un marco de referencia para comprender y emplear operacionalmente el concepto de relevancia, aplicado a las clases de particiones borrosas. Se construyó un criterio que permite discernir sobre la posibilidad de obtener dos particiones borrosas que se presenten como de *alta calidad*. De igual forma, se caracterizaron aquellos operadores de agregación (en particular, reglas recursivas) que permiten evaluar la propiedad de relevancia bajo la propuesta específica, es decir, a partir de la consideración del grupo conmutativo. Lo anterior llevó a establecer la definición de una propiedad posible sobre operadores de agregación: la propiedad de ser *débilmente autodual*.

Finalmente, en la cuarta sección se aplican los resultados de la propuesta al problema de segmentación de imágenes, encontrando en primer lugar que el algoritmo construido proporciona información adicional en términos de la evaluación de una partición borrosa en contraste con varios índices existentes para tal fin.

Lo anterior surge de forma natural, debido a la consideración de varias propiedades para la evaluación de la partición. En concreto, se evalúan propiedades generales como

son el agrupamiento, el solapamiento, los grados de agrupamiento del solapamiento entre otras, y propiedades locales, en términos de la individualidad de cada clase, como lo es la relevancia de las mismas dentro de la partición. Por lo tanto, se confirma que la mejor manera de evaluar una partición borrosa es a través de la consideración integrada de diversas características de la partición y no mediante la evaluación por separado de índices de diversa naturaleza.

En la misma línea, el índice propuesto para determinar un número óptimo de clases, presentó un funcionamiento y resultados consistentes en varias perspectivas. En primer lugar, la determinación del óptimo, en principio, no dependió de los operadores de agregación seleccionados. Para los sistemas de clasificación borrosa abordados, los resultados fueron los mismos. En segundo lugar, ante una coincidencia generalizada por otros índices que evalúan una o dos propiedades, el índice propuesto mantuvo un comportamiento similar, lo cual es un buen indicador de su capacidad para evaluar la partición.

Finalmente, aunque la propuesta consistente desde su desarrollo teórico y, con buenos resultados en la parte práctica, es natural que se requiera de más aplicaciones y usos en diversos contextos para continuar validando la efectividad de la misma. En este sentido, se abren diversas líneas de investigación que se proponen en la última parte del documento.

Líneas de trabajo futuro

La propuesta desarrollada en la presente tesis, ha establecido un amplia línea de investigación de forma particular, en los potenciales de la estructura algebraica construida. El grupo conmutativo propuesto requiere indagar por elementos como la generación de subgrupos borrosos a partir de operadores de agregación, extendiendo la noción de subgrupos borrosos con respecto a una t-norma. Formalmente se tiene: si G es un grupo, una función $\mu : G \to [0, 1]$ se denomina un subgrupo borroso de G con respecto a la t-norma T, si $\forall x, y \in G$, 1) $\mu(x, y) \geq T(\mu(x), \mu(y))$; 2) $\mu(x^{-1}) = \mu(x)$ y 3) $\mu(e) = 1$. Por lo tanto, es natural pensar en extender esta noción a las reglas recursivas dadas por el sistema de clasificación borrosa y estudiar su aplicación sobre particiones borrosas.

De igual forma, la estrecha relación entre subgrupos borrosos y relaciones de similitud invariantes por traslaciones permite indagar en la interpretación de los grupos cociente, los homomorfismos de subgrupos y el núcleo de un grupo en el contexto de tales relaciones.

En particular, sobre los resultados concernientes a la evaluación de particiones borrosas se proponen algunos elementos para ampliar la propuesta. En primer lugar, se requiere un conjunto de imágenes más extenso que permita determinar la efectividad y precisión de la propuesta. En este sentido, se propone una implementación del algoritmo y el número óptimo de clústeres, considerando imágenes pertenecientes a distintos contextos junto con la validación por parte de expertos. En particular, la comparación a partir de imágenes médicas ya clasificadas en estudios reales, se presenta como una alternativa de estudio para la evaluación de la propuesta.

De igual forma, se hace relevante indagar si bajo algunas condiciones, los sistemas de clasificación y el grupo conmutativo permiten evaluar otras propiedades de la partición, por ejemplo la densidad de la partición, la dispersión en cada clase o la

homogeneidad de las clases, entre otras.

Por otro lado, puede llegar a tener gran interés y aplicabilidad considerar operadores de agregación no conmutativos, como puede llegar a presentarse en el trabajo sobre grafos en los cuales los nodos están conectados en un solo sentido. Incluso se puede pensar en introducir las *estructuras pareadas* de acuerdo con [82]. En este sentido, a pesar del aparente aumento en las dificultades de cálculo debido a la información adicional que se incluye, la existencia de opuestos debería ayudar a recopilar más evidencia para evaluar adecuadamente la calidad de una partición y, si es posible, mejorar la confianza en el modelo.

Otra línea de trabajo necesaria a partir de lo desarrollado corresponde con un estudio más profundo sobre los operadores de agregación o reglas recursivas que generan mejores resultados o incluso que deben, de forma exclusiva, ser empleadas en el calculo del número óptimo de clases en una partición borrosa. En este sentido, se hace pertinente evaluar los efectos y resultados de ampliar el conjunto de relaciones de similitud y funciones de disimilitud en los desarrollos planteados.

En la misma línea, durante el desarrollo del Algoritmo 1 se ha establecido como proceso iniciar en una partición con dos clases e ir aumentando el número de clases hasta obtener una partición de *alta calidad*. Se propone en este sentido, realizar un proceso cambiando el orden del funcionamiento. Es decir, iniciar en una partición con un número de clases igual al cardinal del conjunto de objetos, e ir eliminando clases hasta obtener una partición de *alta calidad*.

De igual forma, se sugiere con el fin de ampliar los hallazgos presentados, aplicar la propuesta a otros problemas de clasificación diferente a la segmentación de imágenes, seguramente en otros contextos se establezcan nuevas ventajas y requerimientos adicionales a nuestra propuesta.

Finalmente, una línea de trabajo futura consistirá en el estudio de las raíces de los polinomios característicos de las matrices de disimilitud (y equivalentemente, de las

matrices de similitud), es decir de los valores propios de la matriz y, en particular, de los vectores propios. El interés radica en que son precisamente estos últimos los que permanecen invariantes bajo transformaciones y, por lo tanto, se abre un espectro al trabajo con operadores de agregación invariantes por transformaciones lineales positivas o traslaciones en el sentido propuesto por [45], es decir, operadores tales que $A(px_1 + t, ..., px_n + t) = pA(x_1, ..., x_n) + t$.

Bibliografía

[1] S. Abe. Fuzzy support vector machines for multilabel classification. *Pattern Recognition*, 48(6):2110–2117, 2015.

[2] S. Abe and T. Ruck. A fuzzy classifier with ellipsoidal regions. *IEEE Transactions on Fuzzy Systems*, 5(3):358–368, 1997.

[3] R Alcalá, J Alcalá-Fdez, M. J. Gacto, and F. Herrera. Genetic learning of membership functions for mining fuzzy association rules. *IEEE International Conference on Fuzzy Systems*, 2007.

[4] A. Amo, D. Gomez, Javier Montero, and G. Biging. Relevance and Recundancy in fuzzy classification systems. *Mathware & Soft Computing*, 8(3):203–216, 2001.

[5] A. Amo, J. Montero, G. Biging, and V. Cutello. Fuzzy classification systems. *European Journal of Operational Research*, 156(2):495–507, 2004.

[6] G. Ball and D. Hall. ISODATA, a novel method of data analysis and pattern classification. *Analysis*, (AD699616):1–79, 1965.

[7] A. Baraldi and P. Blonda. A survey of fuzzy clustering algorithms for pattern recognition - Part I, 1999.

[8] A. Baraldi and P. Blonda. A survey of fuzzy clustering algorithms for pattern recognition - Part II. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 29(6):786–801, 1999.

[9] R. Bellman, R. Kalaba, and L. Zadeh. Abstraction and pattern classification. *Journal of Mathematical Analysis and Applications*, 13(1):1–7, 1966.

[10] J. Benzecri. Statistical Analysis As a Tool To Make Patterns Emerge From Data. In *Methodologies of Pattern Recognition*, pages 35–74. 1969.

[11] I. Berget, B. Mevik, and T. Næs. New modifications and applications of fuzzy C-means methodology, 2008.

[12] J. C. Bezdek. Cluster validity with fuzzy sets. *Journal of Cybernetics*, 3(3):58–73, 1973.

[13] J. C. Bezdek. Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology*, 1(1):57–71, 1974.

[14] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. 1981.

[15] J. C. Bezdek and issn = 01650114 journal = Fuzzy Sets and Systems keywords = Cluster analysis,Convex decompositions,Fuzzy relations,Pseudo-metrics,Transitivity mendeley-groups = Bibliografia title = Fuzzy partitions and relations; an axiomatic basis for clustering year = 1978 Douglas H., doi = 10.1016/0165-0114(78)90012-X.

[16] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. Pal. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing (The Handbooks of Fuzzy Sets)*. 1999.

[17] H. Bustince, E. Barrenechea, M. Pagola, and J. Fernandez. The notions of overlap and grouping functions. In *Studies in Fuzziness and Soft Computing*, volume 336, pages 137–156. 2016.

[18] H. Bustince, J. Fernandez, R. Mesiar, J. Montero, and R. Orduna. Overlap functions. *Nonlinear Analysis, Theory, Methods and Applications*, 72(3-4):1488–1499, 2010.

[19] H. Bustince, M. Pagola, R. Mesiar, E. Hüllermeier, and F. Herrera. Grouping, overlap, and generalized bientropic functions for fuzzy modeling of pairwise comparisons. *IEEE Transactions on Fuzzy Systems*, 2012.

[20] L. Cai and H. Kwan. Fuzzy classifications using fuzzy inference networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 28(3):334–347, 1998.

[21] T. Calvo, A. Kolesárová, M. Komorníková, and R. Mesiar. Aggregation Operators: Properties, Classes and Construction Methods. pages 3–104. 2002.

[22] R. Campello and E. R. Hruschka. A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems*, 157(21):2858–2875, 2006.

[23] F. Castiblanco, C. Franco, J. T. Rodríguez, and J. Montero. Degree of global covering and global overlapping in solvency fuzzy classification. In Blanco Mesa F. Gil Lafuente A.M. Merigó J.M. Kacprzyk J León Castro, E., editor, *Intelligent and Complex Systems in Economics and Business*, volume Pre Press. Springer Nature Switzerland AG, 2020.

[24] F. Castiblanco, C Franco, J. T. Rodríguez, and J. Montero. Evaluation of the quality and relevance of a fuzzy partition. *Journal of Intelligent Fuzzy Systems*, (Pre-press):1–16, 2020.

[25] F. Castiblanco, C. Franco, J. T. Rodríguez, and J. Montero. Optimal number of classes in fuzzy partitions. In B. Bede, M. Cock, M. Ceberio, and V. Wu, Y. Kreinovich, editors, *Fuzzy Techniques: Theory and Applications. NAFIPS 2020. Advances in Intelligent Systems and Computing*, volume Pre Press. Springer Nature Switzerland AG, 2020.

[26] F. Castiblanco, C. Franco, J. Rodríguez, and J. Montero. Aggregation operators to evaluate the relevance of classes in a fuzzy partition. In Kearfott R., Batyrshin

I., Reformat M., Ceberio M., and Kreinovich V., editors, *Fuzzy Techniques: Theory and Applications. IFSA/NAFIPS 2019. Advances in Intelligent Systems and Computing*, volume 1000. Springer Nature Switzerland AG, 2019.

[27] F. Castiblanco, C. Franco, J. T. Rodríguez, and J. Montero. Relevance of classes in a fuzzy partition. a study from a group of aggregation operators. In Coelho R. Barreto G., editor, *Fuzzy Information Processing. NAFIPS 2018. Communications in Computer and Information Science*, volume 831. Springer Nature Switzerland AG, 2018.

[28] F. Castiblanco, J. Montero, J. T. Rodríguez, and D. Gómez. Aggregation tools for the evaluation of classifications. In *IFSA-SCIS 2017 - Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems*, pages 1–5, Otsu, Japan, 2017. IEEE.

[29] F. Castiblanco, J. Montero, J. T. Rodríguez, and D. Gómez. Quality assessment of fuzzy classification: An application to solvency analysis. *Fuzzy Economic Review*, 22(1):19–31, 2017.

[30] C. Chen. Design of PSO-based fuzzy classification systems. *Tamkang Journal of Science and Engineering*, 9(1):63–70, 2006.

[31] Y Chen and J. Wang. Support Vector Learning for Fuzzy Rule-Based Classification Systems. *IEEE Transactions on Fuzzy Systems*, 11(6):716–728, 2003.

[32] I. Jen Chiang and Jane Yung Jen Hsu. Fuzzy classification trees for data analysis. *Fuzzy Sets and Systems*, 130(1):87–99, 2002.

[33] O. Cordón, M. Del Jesus, and F. Herrera. A proposal on reasoning methods in fuzzy rule-based classification systems. *International Journal of Approximate Reasoning*, 1999.

[34] V. Cutello and J. Montero. Recursive connective rules. *International Journal of Intelligent Systems*, 14(1):3–20, 1999.

[35] R. Dave. Validating fuzzy partitions obtained through c-shells clustering. *Pattern Recognition Letters*, 17(6):613–623, 1996.

[36] A. Derek, J. Bezdek, M. Popescu, and J. Keller. Comparing fuzzy, probabilistic, and possibilistic partitions. *IEEE Transactions on Fuzzy Systems*, 18(5):906–918, 2010.

[37] J. Dombi. A general class of fuzzy operators, the demorgan class of fuzzy operators and fuzziness measures induced by fuzzy operators. *Fuzzy Sets and Systems*, 8(2):149–163, 1982.

[38] J. Dombi. Basic concepts for a theory of evaluation: The aggregative operator. *European Journal of Operational Research*, 10(3):282–293, 1982.

[39] D. Dubois and H. Prade. A review of fuzzy set aggregation connectives. *Information Sciences*, 36(1-2):85–121, 1985.

[40] R. Duda, P. Hart, and D. G. Stork. Pattern classification. *New York: John Wiley, Section*, 2001.

[41] J. C. Dunn. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57, 1973.

[42] J. C. Dunn. Some recent investigations of a new fuzzy partitioning algorithm and its application to pattern classification problems. *Journal of Cybernetics*, 4(2):1–15, 1974.

[43] J. C. Dunn. Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, 4(1):95–104, 1974.

[44] M. Ferraro and P. Giordani. A review and proposal of (fuzzy) clustering for nonlinearly separable data. *International Journal of Approximate Reasoning*, 115:13–31, 2019.

[45] J. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. 1994.

[46] I. Gath and A. B. Geva. Unsupervised Optimal Fuzzy Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):773–780, 1989.

[47] I. Gitman and M. Levine. An Algorithm for Detecting Unimodal Fuzzy Sets and its Application as a Clustering Technique. *IEEE Transactions on Computers*, C-19(7):583–593, 1970.

[48] D. Gómez, J. T. Rodríguez, J. Montero, H. Bustince, and E. Barrenechea. N-Dimensional overlap functions. *Fuzzy Sets and Systems*, 287:57–75, 2016.

[49] A. Gonzalez and R. Perez. Completeness and consistency conditions for learning fuzzy rules. *Fuzzy Sets and Systems*, 96(1):37–51, 1998.

[50] D. E. Gustafson and W. C. Kessel. Fuzzy Clustering With a Fuzzy Covariance Matrix. In *Proceedings of the IEEE Conference on Decision and Control*, pages 761–766, 1978.

[51] E. Hullermeier, M. Rifqi, S. Henzgen, and R. Senge. Comparing fuzzy partitions: A generalization of the rand index and related measures. *IEEE Transactions on Fuzzy Systems*, 20(3):546–556, 2012.

[52] E. Hüllermeier and S. Vanderlooy. Why fuzzy decision trees are dood rankers. *IEEE Transactions on Fuzzy Systems*, 17(6):1233–1244, 2009.

[53] H. Ishibuchi, T. Murata, and I. B. Türkşen. Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems. *Fuzzy Sets and Systems*, 89(2):135–150, 1997.

[54] H. Ishibuchi and T. Nakashima. Effect of rule weights in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*, 9(4):506–515, 2001.

[55] H. Ishibuchi, K. Nozaki, and H. Tanaka. Distributed representation of fuzzy rules and its application to pattern classification. *Fuzzy Sets and Systems*, 52(1):21–32, 1992.

[56] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. In *ACM Computing Surveys*, volume 31, pages 264–323, 1999.

[57] M. Ju, G Cheng, W. Yixin, An Xingle, Gao Jiantao, Yu Ziqi, Zhang Minqing, Liu Xin, Deng Xueyuan, Cao Shucheng, Wei Hao, Mei Sen, Yang Xiaoyu, Nie Ziwei, Li Chen, Tian Lu, Zhu Yuntao, Zhu Qiongjie, Dong Guoqiang, and He Jian. COVID-19 CT Lung and Infection Segmentation Dataset. 10.5281/zenodo.3757476. https://doi.org/10.5281/zenodo.3757476, April 2020.

[58] M. Jun, W. Yixin, A Xingle, G. Cheng, Y. Ziqi, C. Jianan, Z. Qiongjie, D. Guoqiang, H. Jian, H. Zhiqiang, N. Ziwei, and Y. Xiaoping. Towards efficient covid-19 ct annotation: A benchmark for lung and infection segmentation. *arXiv preprint arXiv:2004.12537*, 2020.

[59] A. Kandel. *Fuzzy Techniques In Pattern Recognition by Abraham Kandel. New York. 1982.* Jhon Wileys & Sons, New York, jul 1982.

[60] M. Kaufmann, A. Meier, and K. Stoffel. IFC-Filter: Membership function generation for inductive fuzzy classification. *Expert Systems with Applications*, 42(21):8369–8379, 2015.

[61] E. Klement and B. Moser. On the redundancy of fuzzy partitions. *Fuzzy Sets and Systems*, 85(2):195–201, 1997.

[62] G. Klir and B. Yuan. *Fuzzy sets and fuzzy logic: Theory and applications*. Upper Saddle River, New Jersey : Prentice Hall, D.L., 1995.

[63] R. Krishnapuram. Generation of membership functions via possibilistic clustering. In *IEEE International Conference on Fuzzy Systems*, volume 2, pages 902–908, 1994.

[64] R. Krishnapuram, H. Frigui, and O. Nasraoui. Fuzzy and Possibilistic Shell Clustering Algorithms and Their Application to Boundary Detection and Surface Approximation—Part I. *IEEE Transactions on Fuzzy Systems*, 3(1):29–43, 1995.

[65] R. Krishnapuram, H. Frigui, and O. Nasraoui. Fuzzy and Possibilistic Shell Clustering Algorithms and Their Application to Boundary Detection and Surface Approximation—Part II. *IEEE Transactions on Fuzzy Systems*, 3(1):44–60, 1995.

[66] R. Krishnapuram and J. Keller. A Possibilistic Approach to Clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98–110, 1993.

[67] R. Krishnapuram and J. Keller. The possibilistic C-means algorithm: Insights and recommendations. *IEEE Transactions on Fuzzy Systems*, 4(3):385–393, 1996.

[68] S. Kundu. Membership functions for a fuzzy group from similarity relations. *Fuzzy Sets and Systems*, 101(3):391–402, 1999.

[69] S. H. Kwon. Cluster validity index for fuzzy clustering. *Electronics Letters*, 34(22):2176–2177, 1998.

[70] J. M. Leski. Generalized Weighted Conditional Fuzzy Clustering. *IEEE Transactions on Fuzzy Systems*, 11(6):709–715, 2003.

[71] S. T. Li and C. C. Chen. A Regularized Monotonic Fuzzy Support Vector Machine Model for Data Mining With Prior Knowledge. *IEEE Transactions on Fuzzy Systems*, 23(5):1713–1727, 2015.

[72] C. Lin and S. Wang. Fuzzy support vector machines. *IEEE Transactions on Neural Networks*, 13(2):464–471, 2002.

[73] D. Lin. An Information-Theoretic Definition of Similarity. In *Icml*, pages 296–304, 1998.

[74] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–296, 1967.

[75] M. Makrehchi and M. Kamel. An information theoretic approach to generating membership functions from real data. In *Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS*, volume 2003-Janua, pages 44–49, 2003.

[76] D. Mandal, C. A. Murthy, and Sankar K. Pal. Formulation of a Multivalued Recognition System. *IEEE Transactions on Systems, Man and Cybernetics*, 22(4):607–620, 1992.

[77] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Proceedings of the IEEE International Conference on Computer Vision*, 2:416–423, 2001.

[78] P. Matsakis, S. Andréfouët, and P. Capolsini. Evaluation of fuzzy partitions. *Remote Sensing of Environment*, 74(3):516–533, 2000.

[79] Andreas Meier, Günter Schindler, and Nicolas Werro. Fuzzy Classification on Relational Databases. In *Handbook of Research on Fuzzy Information Processing in Databases*, pages 586–614. 2011.

[80] S. Mitra and S. K. Pal. Self-organizing neural network as a fuzzy classifier. *IEEE Transactions on Systems, Man, and Cybernetics*, 24(3):385–399, 1994.

[81] S. Mitra and S. K. Pal. Fuzzy sets in pattern recognition and machine intelligence. *Fuzzy Sets and Systems*, 156(3):381–386, 2005.

[82] J. Montero, H. Bustince, C. Franco, J. T. Rodríguez, D. Gómez, M. Pagola, J. Fernández, and E. Barrenechea. Paired structures in knowledge representation. *Knowledge-Based Systems*, 100:50–58, 2016.

[83] J. Mordeson, K. Bhutani, and A. Rosenfeld. *Fuzzy Group Theory*, volume 182 of *Studies in Fuzziness and Soft Computing*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[84] J.D. Murray. *Asymptotic analysis*. Springer Verlag New York, 1984.

[85] T. Nakashima, G. Nakai, and H. Ishibuchi. Improving the performance of fuzzy classification systems by membership function learning and feature selection. *IEEE International Conference on Plasma Science*, 1:488–493, 2002.

[86] D. Nebel, M. Kaden, A. Villmann, and T. Villmann. Types of (dis-)similarities and adaptive mixtures thereof for improved classification learning. *Neurocomputing*, 268:42–54, 2017.

[87] J. Orozco. Similarity and dissimilarity concepts in machine learning. *Technical report, LSI-04-9-R*, 2004.

[88] N. R. Pal, K. Pal, and J. C. Bezdek. Mixed c-means clustering model. In *IEEE International Conference on Fuzzy Systems*, volume 1, pages 11–21, 1997.

[89] S. K. Pal and D. P. Mandal. Linguistic recognition system based on approximate reasoning. *Information Sciences*, 61(1-2):135–161, 1992.

[90] A. Palacios, J. Palacios, L. Sánchez, and J. Alcalá-Fdez. Genetic learning of the membership functions for mining fuzzy association rules from low quality data. *Information Sciences*, 295:358–378, 2015.

[91] T. Pavlidis and R. Chang. Fuzzv Decision Tree Algorithms. *IEEE Transactions on Systems, Man and Cybernetics*, 7(1):28–35, 1977.

[92] W. Pedrycz. Fuzzy sets in pattern recognition: Methodology and methods. *Pattern Recognition*, 23(1-2):121–146, 1990.

[93] J. Qiao and B. Hu. On interval additive generators of interval overlap functions and interval grouping functions. *Fuzzy Sets and Systems*, 323:19–55, 2017.

[94] J. Qiao and Bao Q. Hu. On the migrativity of uninorms and nullnorms over overlap and grouping functions. *Fuzzy Sets and Systems*, 346:1–54, 2018.

[95] M. Rezaee, B. P.F. Lelieveldt, and J. H.C. Reiber. A new cluster validity index for the fuzzy c-mean. *Pattern Recognition Letters*, 19(3-4):237–246, 1998.

[96] K. Rojas, D. Gómez, J. Montero, and J.T. Rodríguez. Strictly stable families of aggregation operators. *Fuzzy Sets and Systems*, 228:44–63, 2013.

[97] A. Rosenfeld. Fuzzy groups. *Journal of Mathematical Analysis and Applications*, 35(3):512–517, 1971.

[98] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C):53–65, 1987.

[99] P. J. Rousseeuw, L. Kaufman, and E. Trauwaert. Fuzzy clustering using scatter matrices. *Computational Statistics and Data Analysis*, 23(1):135–151, 1996.

[100] E. H. Ruspini. A new approach to clustering. *Information and Control*, 15(1):22–32, 1969.

[101] E. H. Ruspini. Numerical methods for fuzzy clustering. *Information Sciences*, 2(3):319–350, 1970.

[102] E. H. Ruspini. New experimental results in fuzzy clustering. *Information Sciences*, 6(C):273–284, 1973.

[103] Ammar Shaker, Robin Senge, and Eyke Hüllermeier. Evolving fuzzy pattern trees for binary classification on data streams. In *Information Sciences*, volume 220, pages 34–45, 2013.

[104] T. Silva and L. Zhao. *Machine learning in complex networks*. 2016.

[105] W. Silvert. Symmetric Summation: a Class of Operations on Fuzzy Sets. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-9(10):657–659, 1979.

[106] P. Sobrevilla and E. Montseny. Fuzzy sets in computer vision: an overview. *Mathware & soft computing*, 10(3):71–83, 2003.

[107] D. Sperber and D. Wilson. Précis of Relevance: Communication and Cognition. *Behavioral and Brain Sciences*, 10(04):697, 1987.

[108] A. Suleman. Measuring the congruence of fuzzy partitions in fuzzy c-means clustering. *Applied Soft Computing Journal*, 52:1285–1295, 2017.

[109] S. Tamura, S. Higuchi, and K. Tanaka. Pattern Classification Based on Fuzzy Relations. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-1(1):61–66, 1971.

[110] R. Wang and K. Mei. Analysis of fuzzy membership function generation with unsupervised learning using self-organizing feature map. In *Proceedings - 2010 International Conference on Computational Intelligence and Security, CIS 2010*, pages 515–518, 2010.

[111] W. Wang and Y. Zhang. On fuzzy cluster validity indices. *Fuzzy Sets and Systems*, 158(19):2095–2117, 2007.

[112] D. Wijayasekara and M. Manic. Data driven fuzzy membership function generation for increased understandability. In *IEEE International Conference on Fuzzy Systems*, pages 133–140, 2014.

[113] K. L. Wu and M. S. Yang. A cluster validity index for fuzzy clustering. *Pattern Recognition Letters*, 26(9):1275–1291, 2005.

[114] X. Xie and G. Beni. A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):841–847, 1991.

[115] M. S. Yang. A survey of fuzzy clustering. *Mathematical and Computer Modelling*, 18(11):1–16, 1993.

[116] J. Yu and M. S. Yang. Optimality test for generalized FCM and its application to parameter selection. *IEEE Transactions on Fuzzy Systems*, 13(1):164–176, 2005.

[117] J. Yu and M. S. Yang. A generalized fuzzy clustering regularization model with optimality tests and model complexity analysis. *IEEE Transactions on Fuzzy Systems*, 15(5):904–915, 2007.

[118] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[119] L. A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 3(2):177–200, 1971.