

Representación efectiva de dinámicas fisiológicas mediante fuzzy rough set: una revisión

Diana Alexandra Orrego Metaute
Especialista en Automatización,
Universidad Pontificia Bolivariana
Docente Tiempo Completo, Investigador Grupo MIRP,
Instituto Tecnológico Metropolitano
Medellín, Colombia
dianaorrego@itm.edu.co

Edilson Delgado Trejos
PhD en Ingeniería con línea de investigación en Automática,
Universidad Nacional de Colombia
Académico Investigador, Líder Grupo MIRP,
Instituto Tecnológico Metropolitano
Medellín, Colombia
edilsondelgado@itm.edu.co

Resumen—Los sistemas biomédicos de última generación registran en intervalos cortos de tiempo la dinámica fisiológica mediante grandes bases de datos. La interpretación adecuada de la información difícilmente puede hacerse por la experticia de un sólo médico, por lo tanto la toma de decisiones se basa sólo en algunas variables seleccionadas. La representación efectiva de variables fisiológicas mediante fuzzy rough set tipo 1 puede ser aplicada para caracterizar y extraer la información relevante de la dinámica fisiológica; sin embargo, estas técnicas poseen el problema de la complejidad de sus algoritmos y alto costo computacional; por lo tanto, se requiere aplicar técnicas de fuzzy rough set tipo 2, asociadas a métodos axiomáticos a través de operadores de aproximación difusa baja y alta como conceptos primitivos para generar un sistema de reducción de dimensiones con tendencia a la disminución de costo computacional en aplicaciones de ingeniería biomédica. En este artículo se presenta la revisión del estado del arte sobre representación efectiva de dinámicas fisiológicas mediante fuzzy rough set, con el fin de determinar la capacidad que poseen este tipo de técnicas para ser incluidas en procedimientos automáticos de toma de decisiones que apoyen el concepto clínico de un especialista.

Palabras clave— Conjuntos Difusos/Aproximados, Dinámica Fisiológica, Reducción de Dimensiones, Representación Efectiva, Extracción/Selección de características.

Abstract— The latest generation of biomedical systems record at short time intervals the physiological dynamic in large databases. The correct interpretation of the information is difficult to obtain by the expertise of a single physician, so the decision is based only on some selected variables. Effective representation of physiological variables by fuzzy Rough Set type 1 can be applied to characterize and extract relevant information from physiological dynamics, however the disadvantages of these techniques are the complexity of their algorithms and the high computational cost, therefore it is necessary to apply fuzzy rough set type 2 techniques, associated with axiomatic methods through low and high diffuse approximation operators as primitive concepts for generating a dimension reduction system with a tendency to lower computational cost in biomedical

engineering applications. This article reviews the state of the art of effective representation of physiological dynamics using fuzzy rough set, in order to determine the ability of these techniques to be included in automatic decision making procedures that support the clinical opinion of a specialist.

Keywords— Fuzzy/Rough Sets, Physiological Dynamics, Dimensionality Reduction, Intrinsic Representation, Feature Extraction/Selection.

I. INTRODUCCIÓN

Los prototipos de última generación en biomédica permiten mostrar en línea una cantidad enorme de datos por paciente que describen o simulan los múltiples procesos fisiológicos del cuerpo humano, con el propósito de ser una herramienta para el diagnóstico y tratamiento médico [1]. Todos estos procesos son fenómenos complejos, que se acompañan o manifiestan mediante señales que reflejan su naturaleza y actividad, pueden ser de diversos tipos: desde señales bioquímicas, como hormonas y neurotransmisores, hasta registros de señales bioeléctricas, como electroencefalografía (EEG), fonocardiografía (FCG) o electrocardiografía (ECG), o mediciones biofísicas como presión arterial y temperatura, entre otros [2]. La caracterización de la dinámica fisiológica como señales normales o patológicas ha permitido el diseño de sistemas expertos de diagnóstico que soportan la decisión médica, debido a que generalmente la decisión está basada sólo en la experticia del especialista [3]. Sin embargo, en la mayoría de los procesos de caracterización de la dinámica fisiológica se contemplan amplios conjuntos de características que conllevan al empleo de grandes recursos computacionales, así como en los posteriores de procesamiento y clasificación de los datos [4].

La universalidad de las bases de datos fisiológicas de alta dimensión ha ayudado al desarrollo de importantes investigaciones en el análisis de características, para obtener conjuntos de variables de representación que aportan la información relevante y redundante del sistema en relación con la inferencia de estados fisiológicos funcionales [5]. En el aprendizaje estadístico, el espacio de representación característico es definido como una dimensionalidad que contiene todos los posibles valores que pueda tomar el patrón, representado por un vector aleatorio, el cual puede ser visto como un punto n -dimensional, con el objetivo de representar el conjunto total de señales de entrenamiento en un espacio donde alguna métrica minimice la distancia entre patrones de una misma clase y maximice la distancia entre los de diferente clase [6]. En este contexto, la reducción de dimensiones busca encontrar la mejor representación de patrones de alta dimensión, que permita obtener un conjunto alterno y compacto con la menor dimensión posible, optimizando una función de evaluación específica en el espacio resultante y regida por una medida de relevancia (estadística, geométrica, informativa o discriminante) encargada de dirigir el contexto de representación [7]. En particular, la reducción de dimensiones puede obtenerse por diferentes técnicas que generalmente se han agrupado en selección y extracción de características [8]. En este sentido, la teoría de conjuntos difusos propuesta por Zadeh [9] en 1965 (FST), la cual es una extensión de la noción de conjuntos clásica para modelar la incertidumbre en términos de clasificación de membresía, permite evaluar los subconjuntos usando una función, o métricas de evaluación, con el fin de seleccionar las más importantes derivadas de las clases de decisión [10]. Un sistema básico difuso incluye elementos como fusificador, reglas, motor de inferencia y defusificador. La teoría de conjuntos Rough (RST) propuesta por Pawlak [11], es un método de selección de características nuevo donde se preserva la semántica de las funciones, lo cual permite analizar los hechos ocultos de los datos sin necesitar información adicional, como umbrales o conocimiento de expertos, para obtener un subconjunto denominado "reducto" con las características originales de mayor información. La aplicación de RST sólo se puede realizar sobre conjuntos de datos con valores de tipo real, por lo tanto es necesario realizar

una discretización previa de los datos, generando como resultado la pérdida de información [9]. La teoría de conjuntos rough, complementada con la teoría de conjuntos difusos, da lugar a un nuevo método de selección de características denominado "Fuzzy Rough Set" (FRS) [12] en el que ambas teorías, a pesar de relacionarse, son diferentes. Esta técnica híbrida genera aproximaciones rough de conjuntos difusos por medio de relaciones de similitud o particiones difusas. Adicionalmente, ofrece alto grado de flexibilidad, soluciones robustas y herramientas avanzadas para la selección de características en el análisis de datos [13]. Actualmente, las investigaciones en el campo de FRS se han centrado en ambientes difusos tipo 1, lo cual no permite modelar la incertidumbre de manera directa ya que se caracteriza por funciones de membresía clásicas, donde se presentan dificultades en la modelación de incertidumbres comunes para la dinámica fisiológica por la falta de fronteras discriminantes en las clases y falta de información [14]. Los FRS tipo 2 introducen intervalos, parámetros y conjuntos difusos para describir la función de membresía incierta del conjunto difuso tipo 1, mediante la adopción de conjuntos rough [15].

Este artículo presenta una revisión sobre espacios de representación efectiva de la dinámica fisiológica logrados mediante fuzzy rough set con el fin de precisar la frontera del estado del arte de temas relacionados con el entrenamiento de procesos automáticos orientado a la reducción de dimensiones en sistemas de soporte de diagnóstico clínico automático.

II. ESPACIOS DE REPRESENTACIÓN CARACTERÍSTICA

Sobre la dinámica fisiológica se debe realizar un proceso de estimación de atributos, o características, que permitan la extracción de la información intrínseca embebida en los estados funcionales determinados por los cambios presentados en las señales que se usan para el diagnóstico de cada caso particular [7]. En algunas ocasiones, el espacio resultante está compuesto por un número alto de dimensiones, lo cual se traduce en una representación fisiológica muy compleja, que a veces puede estar conectada con el sentido clínico o simplemente puede ser tomada como un espacio de representación abstracta (por ejemplo,

coeficientes wavelet, índices fractales, indicadores estadísticos, entre otros) [16].

A. Estimación de parámetros

Al proceso de atribución de una descripción paramétrica a un objeto se denomina *estimación de parámetros*, este proceso es una parte importante del procesamiento o tratamiento de datos y se basa en la regularidad estadística de una gran cantidad de muestras [17]. El grado de confianza del intervalo indica la probabilidad de ocurrencia de los valores estimados en él. Así, en este método de estimación es necesario saber la regularidad de distribución de las muestras, de lo contrario, las hipótesis acerca de ellas y la evaluación del grado de incertidumbre serán erróneas, reduciendo la confiabilidad del método [18].

B. Procesamiento de datos

1) Normalización

En muchas situaciones prácticas un diseño es confrontado con los valores que no expresan lo mismo en ciertos rangos dinámicos [19]. Así, las características con un gran número de valores, pueden tener una gran influencia en el costo de su función de las características con valores pequeños, aunque esto no necesariamente es reflejado en el diseño del clasificador [20]. Un procedimiento comúnmente usado es la normalización estadística, expresada en función de la media y la desviación estándar [7].

2) Remoción de valores atípicos

Los datos atípicos, también conocidos como *outliers*, son las observaciones que parecen tener un comportamiento distinto, o haber sido generados de forma diferente [21]. Cuando existe más de un dato atípico en la muestra, es posible que se presenten efectos de enmascaramiento, en el cual dichas observaciones se ocultan entre sí [22]. Los datos atípicos pueden ser detectados bajo dos diferentes perspectivas, bien sea de manera univariada o multivariada [23]. Calcular medidas robustas de descentralización y dispersión mediante el cálculo de la mediana o la mediana de las desviaciones absolutas con respecto a la mediana, lo cual es una medida robusta de dispersión, permite detectar de forma univariada la presencia de datos atípicos [7]. Adicionalmente, existen diferentes formas de comprobar la homogeneidad de la muestra, la cual

refleja si los datos se separan mucho, o por el contrario se concentran alrededor de la media, mediante el análisis del coeficiente de kurtosis. Este coeficiente permite medir la relación entre la variabilidad de las desviaciones y la desviación media. La detección de datos atípicos de forma multivariada se realiza mediante una técnica de maximizar y minimizar el coeficiente de kurtosis de los datos proyectados [21] y [24].

3) Verificación de la normalidad

Muchas de las técnicas de análisis multivariado están basadas en modelos paramétricos, por lo tanto, existen claras restricciones en cuanto al tipo de distribución a la que dichas variables deben aproximarse [25]. Por otra parte, es también importante asegurar la homogeneidad de la muestra, mediante el análisis de la posible presencia de valores atípicos debidos a errores de medida u otras causas de heterogeneidad [26]. Es necesario realizar la verificación de la normalidad de los datos mediante el juicio sobre la estructura Gaussiana de los datos, a partir del análisis de los histogramas y de la respectiva prueba de hipótesis [27]. La división de valores que toman las variables aleatorias en rangos de alguna vecindad genera diferentes clases de representación, de las cuales la más conocida corresponde al diagrama del número de observaciones en función de la localización de cada rango de la variable aleatoria, conocida como histograma. Sin embargo, la cantidad de intervalos de agrupación dependen del volumen de la muestra [7].

4) Distribución intraclases

Si la prueba de verificación de la distribución da como resultado el rechazo de la hipótesis de normalidad, entonces se deben tomar medidas para transformar la observación de tal manera que pueda cumplirse la aceptación de la hipótesis sobre la normalidad de los datos intraclase [28]. En este sentido el análisis de los histogramas puede ayudar a sugerir alguna forma de acomodación de los datos. Inicialmente se puede partir de un conjunto de posibles transformaciones [29]. En tareas asociadas al proceso de bioseñales, es común el empleo de la distribución logarítmica normal. Después de realizar la transformación, se debe realizar de nuevo la prueba de verificación, se toma aquella transformación que permita aceptar la hipótesis de normalidad intraclase, o bien, aquella que más se aproxime [7].

III. REDUCCIÓN DE DIMENSIONES.

Con el objetivo de reducir dimensiones y así evitar el incremento exponencial del número de muestras fisiológicas necesarias para el entrenamiento de un sistema automático inteligente, es requerida la etapa de extracción/selección de características para obtener el espacio de representación característica [30]. Por lo tanto, es importante tener presente el concepto de “maldición de la dimensionalidad” [7] en el cual se describe el problema que enfrentan los métodos de aprendizaje, donde aumentar el número de características conlleva a un incremento exponencial en el número de objetos de entrenamiento, es decir, a medida que aumenta la dimensionalidad, las características del conjunto de datos de entrada tienden a ser demasiado dispersos para entrenar los algoritmos de aprendizaje de forma eficiente y se requieren muchas más muestras de entrenamiento [31]. Es por esto que las técnicas de reducción de dimensiones permiten transformar conjuntos de datos de grandes dimensiones en conjuntos de dimensiones más pequeñas, para mejorar el procesamiento de datos y revelar la estructura oculta que ayudan en la comprensión y visualización de la dinámica física o fisiológica [32].

En los últimos años ha crecido el interés en las investigaciones relacionadas con reducción de dimensiones para encontrar un método que permita disminuir el número de variables de representación usando la información contenida dentro del conjunto original de los datos y que preserve el significado de las características relevantes. La técnica RST puede ser usada como una herramienta para descubrir la dependencia de los datos y minimizar el número de características contenidas en el conjunto de datos sin utilizar información adicional, pero requiere de una discretización de los datos antes de aplicar el método, ya que el método no diferencia datos de valor discreto de los de valor real [33]. Como resultado se han propuesto extensiones al método RST basados en relaciones de similitud y tolerancia. La técnica FRS (Fuzzy Rough Set) fue propuesta como una extensión del método RST donde se encapsulan los conceptos de vaguedad mediante conjuntos difusos y duplicados, tomando conjuntos rough relacionados pero de diferente valor (discreto o real), los cuales se producen como consecuencia de la incertidumbre en el conocimiento [31].

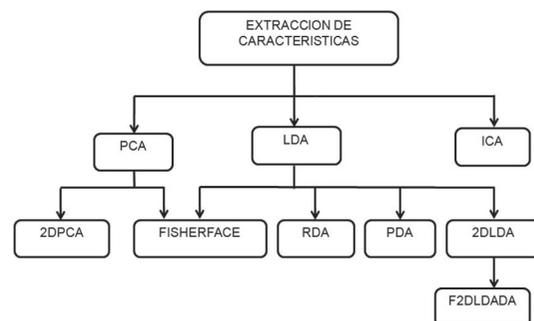
1) Extracción de características

En la extracción de características todas las variables son utilizadas y los datos se transforman de un espacio de dimensión alta a uno de dimensión baja, conservando la mayor parte de información deseada y optimizando la dirección del mapeo [8]. Esta transformación puede ser una combinación de las variables originales de forma lineal o no lineal, supervisado o no supervisado [34]. La extracción de características permite mejorar el ancho de banda de los datos de entrada para un mejor desempeño en el clasificador al proporcionar un conjunto de características relevantes con mínima redundancia [35].

La literatura de extracción de características reporta dos grupos de técnicas: lineales y no lineales, siendo las primeras más comunes y adecuados en la reducción de datos de estructura lineal [36] y [32], como Análisis de Componentes Principales (PCA) [37] y Escalamiento Multidimensional (MDS) [38] mientras las segundas como Embebimiento Lineal Local (LLE) [39] e Isomap (Mapeo Isométrico de Caracter) [40] preservan las medidas de distancia cuando se mapean los datos a un espacio de dimensión reducida. Sin embargo, estos métodos tienden a destruir la semántica fundamental de las características después de la reducción, como pasa en los métodos basados en transformaciones, y en otros casos requieren información adicional acerca del conjunto de datos dado para la umbralización, como pasa en los métodos basados en entropía [41].

De acuerdo con lo expuesto en la Fig. 1, las técnicas lineales comúnmente usadas para extracción de características son [42]: Análisis de Componentes Principales (PCA) [43], Análisis Discriminante Lineal (LDA) [44], Análisis de Componentes independientes (ICA) [45].

FIG. 1. DIAGRAMA EN BLOQUES: EXTRACCIÓN DE CARACTERÍSTICAS



Fuente: Autor del proyecto

PCA y LDA son los métodos más utilizados para extracción de características en la mayoría de aplicaciones de reconocimiento de patrones, sin perder la información de varianza en los datos de entrada [46]. Originalmente LDA fue desarrollado para aprendizaje supervisado, y para una mayor eficacia en la separabilidad de clases, especialmente para encontrar la función óptima discriminante lineal en problemas de clasificación, mientras PCA determina la información relevante para la representación de clases mediante una base ortogonal, lo cual es adecuado para análisis de datos gaussianos, sin embargo, cuando se presentan distribuciones de datos no gaussianas, el método es deficiente. En conclusión, la aplicación de LDA es limitada cuando el número de muestras por clase es pequeño, mientras en PCA se limita cuando el conjunto de entrenamiento presenta un gran número de muestras [47].

A diferencia de LDA y PCA, ICA puede ser considerado como una generalización del estándar PCA, el cual se caracteriza por la independencia estadística sobre los componentes extraídos y no tiene restricción de ortogonalidad. ICA ha sido de gran aplicación para la separación de fuentes estadísticamente independientes y la extracción de características en el análisis de datos de tipo no gaussianos [48]. ICA es también una técnica de reducción de características redundantes que asegura que la información mutua entre los canales de salida filtrada es cero [49]. ICA y PCA siguen siendo modelos lineales, lo que los hace inadecuados para describir datos de distribución compleja y no lineal. Por este motivo, otras técnicas son propuestas en la literatura cuando la naturaleza de los datos es compleja y son afectados por operadores no lineales. Kernel ICA es una técnica que contiene un mapeo de núcleo no lineal apto para características que describen procesos no lineales. Kernel ICA no es la "Kernelización" de un algoritmo ICA existente, sino un método nuevo sobre medidas de dependencia basadas en Kernel [50]. Este método ha sido ampliamente aplicado en procesamiento de imágenes para descubrir los vectores fundamentales de la formación ellas, y son usados como modelos temporales para extracción de características de imágenes de prueba, las cuales son completamente diferentes de las de formación [51]. Otro

método muy popular reportado en la literatura para el reconocimiento facial es denominado Fisherface, el cual usa inicialmente PCA para reducir la dimensión del espacio de características original para $N-c$ (donde N es el número de muestras totales de entrenamiento y c es el número de clases) y luego aplica LDA para reducir la dimensión a d ($d \leq c$). En el proceso de transformación, la componente de proyección más pequeña $c-1$ es descartada, lo que puede implicar pérdida de información discriminante y la etapa basada en PCA no puede garantizar la transformación de la información intraclase, puesto que la matriz de dispersión podría ser no singular [52]. Un método más sistemático propuesto es RDA (Análisis Discriminante Regularizado), el cual trata de obtener estimaciones más confiables mediante la corrección de la distorsión de los valores propios, usando un método de regularización tipo riesgo [53]. PDA (Análisis Discriminante Penalizado) es otra versión regularizada de LDA, cuyo objetivo es, además de superar el problema del tamaño de muestras pequeñas, suavizar los coeficientes de vectores discriminantes que mejoren la interpretación [54].

PCA, LDA, RDA y PDA, clasificados como métodos de una dimensión, no son aptos para escalonamiento en aplicaciones con altas dimensiones, debido a las largas matrices de covarianza con un número relativamente pequeño de muestras de entrenamiento, que dificultan una evaluación más precisa de la matriz [55]. Como una solución a este problema se propone el método de Análisis de Componentes Principales de 2 Dimensiones 2DPCA, basado en matrices bidimensionales en lugar de vectores de una dimensión [56]. A diferencia de PCA, 2DPCA posee una matriz de covarianzas mucho más pequeña, además de la rapidez para calcular los vectores propios derivados de dicha matriz para la caracterización de los datos [57]. Partiendo de los resultados exitosos de la aplicación de 2DPCA, se propuso 2DLDA como un Análisis Discriminante Lineal basado en vectores y en matrices, que permiten evaluar la matriz de covarianzas con precisión y reducir el costo computacional. Sin embargo, 2DLDA asume el mismo nivel de tipicidad de cada muestra para la correspondiente clase [58]. Actualmente se propone el nuevo método F2DLDA (2DLDA Difuso), basado en los criterios discriminantes 2D

fisherface y la teoría de conjuntos difusos, para calcular el grado de membrecía de la matriz utilizando el algoritmo FKNN (k-vecino más cercano difuso) que produce el correspondiente grado de pertenencia a una clase [59].

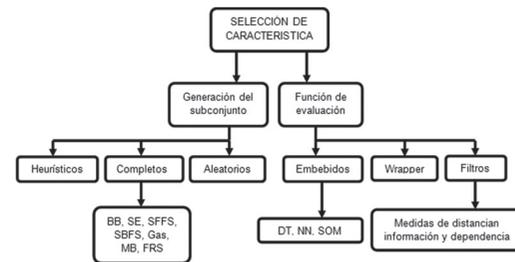
2) Selección de características

La selección de características es usada para identificar y remover las variables que no contribuyen al proceso de clasificación (información redundante e irrelevante) [60]. Este proceso es descrito en [61] como la obtención de un subconjunto de características estimado mediante una función de evaluación, que se compara constantemente con el anterior hasta obtener el mejor de ellos por medio del criterio de parada. Mediante la validación se verifica la calidad que satisface las condiciones del proceso.

Como se describe en la Fig. 2, la selección de características se compone de diferentes familias de técnicas que pueden agruparse en tres métodos de acuerdo con el procedimiento de generación de subconjuntos y a la función de evaluación [62]: completos, que examinan todas las posibles combinaciones de N características (2^N combinaciones), los cuales son muy costosos computacionalmente pero el éxito de estas técnicas es que aseguran el hallazgo del subconjunto óptimo, a diferencia de los heurísticos que hacen una evaluación parcial (N^2 combinaciones), pero mejoran la velocidad de procesamiento ya que el espacio de búsqueda es menor, y los aleatorios que realizan una búsqueda a partir de todo el espacio posible (2^N combinaciones), con la diferencia de que usa una regla aleatoria de acuerdo a unas condiciones iniciales, las cuales podrían encontrar el subconjunto requerido sin procesar todo el espacio de posibilidades, pero el éxito depende fuertemente de las condiciones iniciales. Los resultados de los últimos dos métodos no aseguran la obtención del mejor subconjunto de características, pero sí del más adecuado según los requerimientos establecidos [61]. De acuerdo con la función de evaluación se conocen 3 grupos: embebidos, filtros y wrapper. Los métodos embebidos hacen parte integral de un predictor específico o dependen de la función de evaluación como lo hacen los árboles de decisión (DT) [63], redes neuronales (NN) [64] y mapas auto-organizativos (SOM) [65]. Los filtros [62] son completamente independientes a la función del clasificador, pero requieren de medidas

de evaluación para determinar el mejor subconjunto (e.g., distancia, información, dependencia, consistencia, entre otros), mientras los wrapper [66] para su proceso de selección utilizan como medida la tasa de error del clasificador, incluyendo el costo computacional debido a la estrategia de clasificación usada. En estos tres métodos se obtienen generalmente mejores resultados que en los basados en la generación de subconjuntos, pero el costo computacional sigue siendo representativo [67].

FIG. 2. DIAGRAMA EN BLOQUES: SELECCIÓN DE CARACTERÍSTICAS



Fuente: Autor del proyecto

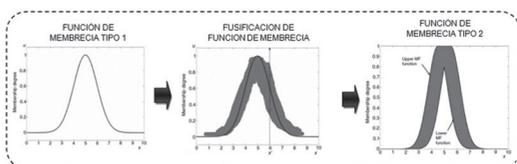
En general, la literatura reporta diferentes técnicas de selección de características que incluyen algoritmos de búsqueda óptima como Branch and Bound y de Búsqueda Exhaustiva [68], sin embargo, son limitados al número de características, ya que el costo computacional aumenta de forma exponencial al número de variables. Las estrategias de búsqueda secuencial como SFS (Sequential Forward Selection) y SBS (Sequential Backward Selection) logran el mejor subconjunto de características de acuerdo con un número de características preestablecido pero no permiten la reelección de ellas [69]. El algoritmo "Plus-L-Minus-R" [70] aplica una búsqueda secuencial más compleja para dar solución a este problema pero se limita por la dificultad de seleccionar los valores adecuados de L y R. A partir de SFS y SBS se evoluciona a las técnicas SFFS (Sequential Forward Floating Selection) y SBFS (Sequential Backward Floating Selection) que permiten la reelección de características incluidas o removidas, mejoran el costo computacional de una forma aceptable [8]. Otras técnicas robustas han sido investigadas recientemente como Algoritmos Genéticos (GAs) [71], Neuro-fuzzy [72], Relief [73], Markov Blanket [74], basados en condicionales de dependencia, independencia y relevancia, pero son algoritmos que

aún no han sido muy desarrollados y presentan un rendimiento muy bajo según datos de la vida real [75].

VI. FUZZY ROUGH SET SOBRE ESPACIOS DE REPRESENTACIÓN

La teoría de conjuntos difusos fue propuesta por Zadeh en 1965 [76], como un método para modelar la incertidumbre en términos de clasificación de membresía, el cual fue nombrado como lógica difusa tipo 1 [77]. Un sistema difuso consta de una interfaz de fusificación para transformar los datos convencionales ingresados por el operador en conjuntos difusos tipo 1, luego un sistema de inferencia usa los conjuntos difusos además de la base del conocimiento para hacer inferencias por medio de un método de razonamiento, y finalmente, una interface de desfusificación traduce la información difusa a la información convencional [78]. La función de membresía tipo 1 es totalmente bivaluada [79], y sus formas convencionales pueden ser triangular, trapezoidal, gaussiana, entre otras, además no permiten caracterizar incertidumbres lingüísticas, es decir, modelar directamente clases de incertidumbre con base en reglas como el lenguaje natural de diferentes significados o medidas tomadas en entornos ruidosos [80], como es el caso del entorno derivado de la dinámica fisiológica. Si para una función de membresía tipo 1 como la de la Fig. 3, se fusifica nuevamente, entonces se obtiene una función de membresía tipo 2, en este caso toma valores diferentes que no son ponderados de la misma función de membresía tipo 1, por lo tanto, se puede asignar una distribución de la amplitud en todos los puntos [78].

FIG. 3. FUNCIÓN DE MEMBRESIA TIPO 2.

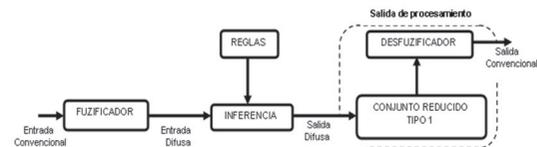


Fuente: Hanbook of Granular Computing [78]

Los sistemas difusos tipo 2, como se describe en la Fig. 4, reducen al mínimo los efectos de la incertidumbre [81], son muy útiles en circunstancias donde es difícil determinar una función de membresía exacta y presentan incertidumbre

en la medición. Los tipo 1 son utilizados recientemente en la industria para aplicaciones de control [82]. La evolución hacia los sistemas difusos tipo 2 ha tomado más tiempo de lo esperado debido al costo computacional que requieren los algoritmos en ambientes de tiempo real [83]. La literatura reciente ha concluido que los sistemas difusos tipo 1 son basados en lógica convencional (numérica, alfanumérica y binaria), mientras los tipo 2 se basan en lógica difusa, que proporcionan un mejor modelo para la incertidumbre inherente al problema y robustez [84].

FIG. 4. SISTEMA DIFUSO TIPO 2.



Fuente: Autor del proyecto

La teoría de conjuntos rough (RST por sus siglas en inglés), inicialmente propuesta por Pawlak en 1982, es un método matemático para el tratamiento de la incertidumbre, vaguedad e imprecisión en el análisis de datos. Su éxito fundamentalmente se debe a la exploración de patrones ocultos embebidos en la base de datos sin demandar información adicional como umbrales o conocimiento de expertos. Dado un conjunto de datos con valores de atributos discretizados (entendiéndose como discretizados a la traducción de valores reales a nominales o simbólicos), la RST permite encontrar un subconjunto dentro de los atributos originales que aportan la mayor información. Sin embargo, presentan limitaciones en el manejo de características ubicadas en la escala de los reales [85]. Convertir a difuso los valores reales y crear un nuevo conjunto de datos con valores simbólicos antes de aplicar RST como método para reducir dimensiones ha sido una solución a dicha limitación, pero presenta considerables pérdidas de información [86].

En los últimos años, las investigaciones en FRS (*Fuzzy Rough Set*) se han centrado principalmente sobre todas las derivaciones que pueden surgir sobre conjuntos difusos. En [87] se propone un método que integra un algoritmo de inducción de reglas difusas denominado Quick Reduct, donde mediante técnicas FRS se realiza la selección de características mediante la función de dependencia. Sin embargo, el algoritmo no es convergente en muchas bases de datos reales debido a su po-

bre criterio de finalización y el crecimiento exponencial de la complejidad computacional [88]. En [89], Tsang et al., establecen las bases sólidas de la matemática para la reducción de características FRS. Adicionalmente, se propone un algoritmo basado en la matriz de discernibilidad para calcular todos los posibles subconjuntos de características relevantes, pero la validación es muy restringida y no se reportan resultados sobre bases de datos reales, como son las bases de datos de funcionamiento fisiológico. En [90], Maji et al., proponen un método basado en FRS usando el concepto de f -información en espacios de aproximaciones difusas, encontrando las características no redundantes y relevantes de conjuntos de datos en la escala de los reales. En [85] Wu et al., proponen el concepto de conjuntos *fuzzy rough* generalizado a entornos *fuzzy* tipo 2 para la reducción de características. Adicionalmente, desarrollaron el algoritmo “IT2 fuzzy-rough QuickReduct”, el cual utiliza la función de dependencia para seleccionar las características que se pueden agregar al candidato de reducción actual. El algoritmo termina cuando la adición de cualquier característica restante no aumenta el grado de dependencia del *fuzzy rough* tipo 2. La demostración fue lograda mediante un ejemplo de un sistema de información donde la reducción de características genera diferentes resultados cuando se compara con *fuzzy rough set* convencional, donde el grado de dependencia disminuye para *fuzzy rough* tipo 2 por el uso de reglas condicionales sobre el amplio conjunto de características.

V. CONCLUSIONES

La dinámica fisiológica representada mediante señales biomédicas (e.g. ECG, FCG, EEG, EMG, entre otros) son de alta complejidad y al momento de caracterizar su dinámica se requiere de espacios de representación de alta dimensión, lo que implica el uso de grandes recursos computacionales para su procesamiento, almacenamiento y transmisión. Aunque los sistemas modernos de cómputo, debido a su alto rendimiento, permiten analizar cantidades grandes de datos en espacios de dimensiones altas, cuando el número de características se incrementa, resultan inconvenientes que requieren ser tratados para asegurar el buen rendimiento del sistema automático. El aumento en el número de dimensiones del espacio de ca-

racterísticas es originado, entre otras cosas, por las exigencias de calidad y precisión en la representación de los estados funcionales, teniendo en cuenta que la complejidad crece exponencialmente a medida que se adicionan dimensiones; de acuerdo con esto, el problema se relaciona con el aumento significativo del tiempo de ejecución de los algoritmos y el exagerado costo en los estudios. En este sentido, reducir el número de dimensiones del espacio de representación, a través de un método basado en la extracción y selección de características, que contenga información de expertos mediante la inclusión de conjuntos difusos para distinguir aquellas variables que ofrezcan mayor capacidad discriminante, sin afectar negativamente la precisión de clasificación, puede llegar a reducir hasta en un amplio porcentaje los costos de procesamiento y en general costos económicos a sectores hospitalarios o de ámbito industrial que usan este tipo de aplicaciones.

Las teorías de conjuntos *rough* y conjuntos *fuzzy* han demostrado tener un buen funcionamiento cuando se crea un híbrido entre ellas, de forma que se aprovechan las bondades inherentes de ambas técnicas, surgiendo el método Fuzzy Rough Set (FRS). Esta nueva técnica proporciona medios de reducción de dimensiones para valores simbólicos y reales utilizando medidas de similaridad. Los conjuntos *fuzzy rough* reúnen los conceptos relacionados pero distintos de imprecisión e indistinguibilidad debidas al ruido provocado por la incertidumbre natural del conocimiento.

En este artículo se presentó una revisión del estado del arte basado en el estudio de la hibridación de los métodos *fuzzy set* y *rough set* para la selección de características con el objetivo de modelar la vaguedad y duplicidad como resultado de la incertidumbre en el conocimiento de la dinámica fisiológica. Aunque los algoritmos de reducción de dimensiones basados en técnicas *fuzzy rough set* tipo 2 poseen el problema de ser altamente costosos desde el punto de vista computacional y son difícilmente traducibles a un lenguaje nominal o simbólico, se puede obtener mediante técnicas de *fuzzy rough sets* tipo 2, asociadas a métodos axiomáticos a través de operadores de aproximación *fuzzy* baja y alta como conceptos primitivos, para generar un

sistema de reducción de dimensiones con tendencia a la disminución de costo computacional para aplicaciones en ingeniería biomédica.

REFERENCIAS

- [1] A. López and C. Vázquez, "Algoritmo de compresión para la señal de pulso arterial periférico," *Universidad, Ciencia y Tecnología*, vol. 9, no. 34, pp. 75-79, Junio 2005.
- [2] D. Sánchez, "Procesado y transmisión de señales biomédicas para el diagnóstico de trastornos y enfermedades del sueño," *Universidad de Cádiz, Tesis Doctoral* 2008.
- [3] M. Imhoff, R. Fried, U. Gather, and V. Lanius, "Dimension Reduction for Physiological Variables Using Graphical Modeling," in *AMIA 2003 Symposium Proceedings, Germany, 2003*, pp. 313-317.
- [4] J. H. Rivera, G. Castellanos, and J. S. Mejía, "Selección efectiva de características para bioseñales utilizando el análisis de componentes principales," *Scientia et Technica*, vol. 13, no. 34, pp. 127-131, May 2007.
- [5] L. Xie and J. Li, "A Novel Feature Extraction Method Assembled with PCA and ICA for Network Intrusion Detection," in *International Forum on Computer Science-Technology and Applications*, vol. 3, Chongqing, 2009, pp. 31-34, 25-27.
- [6] K. Hongfa and C. Yongguang, "An Interpolation Technique Based on Grey Relational Theory," *The Journal of Grey System*, vol. 18, no. 1, pp. 79-83, 2006.
- [7] E. Delgado, G. Castellanos, and M. Vallverdú, *Análisis de Relevancia en Espacios de Representación Orientado al Soporte de Diagnóstico: Aplicaciones en la Dinámica Cardíaca*. Manizales: Ediciones Universidad Nacional de Colombia, 2009, vol. I.
- [8] L. Zhang, Y. Zhong, B. Huang, J. Gong, and P. Li, "Dimensionality Reduction Based on Clonal Selection for Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4172-4186, Dec 2007.
- [9] L. A. Zadeh, "FuzzySets," *Information and Control*, vol. 8, pp. 338-353, 1965.
- [10] L. Zadeh, "Probability measures of fuzzy events," *J. Math. Anal. Appl.*, vol. 23, pp. 421-427, 1968.
- [11] Z. Pawlak, *Rough Sets, Theoretical Aspects of Reasoning About Data*. Netherlands: Kluwer Academic Publishers, 1991.
- [12] Y. Y. Yao, "A Comparative Study of Fuzzy Sets and Rough Sets," *Information Sciences*, vol. 109, no. 1-4, pp. 227-242, 1998.
- [13] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *International Journal of General Systems*, vol. 17, no. 2 y 3, pp. 191-209, 1990.
- [14] D. Dubois and H. Prade, *Fuzzy sets and systems: theory and applications*, Mathematics in Science and engineering. New York: ACADEMIC PRESS, INC, 1980, vol. 144.
- [15] D. Chen and S. Zhao, "Local reduction of decision system with fuzzy rough sets," *Fuzzy Sets and Systems*, vol. 161, no. 13, pp. 1871-1883, Jul. 2010.
- [16] E. Delgado, F. A. Sepúlveda, S. Röthlisberger, and G. Castellanos, "The Rademacher Complexity Model over Acoustic Features for Improving Robustness in Hypernasal Speech Detection. Volume. Editors: Ni," in *Computers and Simulation in Modern Science*. UK: WSEAS Press, University of Cambridge, 2011, pp. 130-135.
- [17] T. Wagner, "Nonparametric estimates of probability densities," *Information Theory, IEEE Transactions on*, vol. 21, no. 4, pp. 438-440, Jul 1975.
- [18] B. Xia, D. Hong, and K. Hongfa, "Study on Grey Parameter Estimation Approach of Small Samples," in *Grey Systems and Intelligent Services, 2009. GISIS 2009.*, Nov.2009, pp. 311-315.
- [19] D. C. Caccia, D. Percival, M. J. Cannon, R. Gary, and J. B. Bassingthwaighte, "Analyzing exact fractal time series: evaluating dispersional analysis and rescaled range methods," *Physica A: Statistical and Theoretical Physics*, vol. 246, no. 3-4, pp. 609-632, December 1997.
- [20] J. B. Bassingthwaighte and G. M. Raymond, "Evaluating rescaled range analysis for time series," *Annals of Biomedical Engineering*, vol. 22, no. 4, pp. 432-444, July 1994.
- [21] D. Peña and F. J. Prieto, "Multivariate Outlier Detection and Robust Covariance Matrix Estimation," *Technometrics*, vol. 43, no. 3, pp. 286-310, Aug. 2001.
- [22] A. C. Rencher, *Methods of Multivariate Analysis*. , 2nd ed. New York: Wiley Series in Probability and Statistics, 2002.
- [23] D. Peña, *Análisis de datos multivariantes.*: Mc Graw Hill, 2002.
- [24] Z. M. Bokal and R.B. Sinitzyn, "Nonparametric method for estimating spoken language sound multivariate probability density function," in *Microwaves, Radar and Remote Sensing Symposium, 2008. MRRS 2008*, Kiev , Sept. 2008, pp. 170-171, 22-24.
- [25] V. Babovic, M. Keijzer, and M. Stefansson, "Chaos theory, optimal embedding and evolutionary algorithms," *Tech. Rep. TR-9800463*, Danish Technical Research Council (STVF), January 2001.

- [26] M. L. Petrovich and M. I. Davidovich, *Estimación Estadística y Prueba de Hipótesis en el Computador (Rus)*. Moscú: Finansy i Statistika, 1989.
- [27] L. Baringhaus and N. Henze, "Limit Distributions for Measures of Multivariate Skewness and Kurtosis Based on Projections," *Journal of Multivariate Analysis*, vol. 38, no. 1, pp. 51–69, July 1991.
- [28] K. Umayahara and Y. Nakamori, "N-dimensional views in fuzzy data analysis," in *Intelligent Processing Systems 1997. ICIPS '97. 1997 IEEE International Conference on*, vol. 1, Oct. 1997., pp. 54-57.
- [29] T. W. Anderson and Bahadur R. R., "Classification Into Two Multivariate Normal Distributions With Different Covariance Matrices," *The Annals of Mathematical Statistics*, vol. 33, no. 2, pp. 420–431, 1962.
- [30] E. Delgado, A. Perera, M. Vallverdú, P. Caminal, and G. Castellanos, "Dimensionality reduction oriented toward the feature visualization for ischemia detection," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 590-598, Mar. 2009.
- [31] R. Jensen and Q. Shen, "Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches," *IEEE Transactions On Knowledge and Data Engineering*, vol. 16, no. 12, pp. 1457-1471, Dec. 2004.
- [32] F. S. Tsai and K. L. Chan, "Dimensionality Reduction Techniques for Data Exploration," in *Information, Communications & Signal Processing, 2007 6th International Conference on*, Dec. 2007, pp. 1-5, 10-13.
- [33] G. Hu and S. Yuemei, "An Attribute Reduction Method based on Fuzzy-Rough Sets Theories," in *First International Workshop on Education Technology and Computer Science*, vol. 3, China, Mar. 2009., pp. 828-831.
- [34] I.K Sethi, "Entropy nets: from decision trees to neural networks," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1605-1613, Oct. 1990.
- [35] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction Foundations and Applications*. Netherlands: Springer-Verlag, 2006.
- [36] D. Gering. (2002., Apr.) *Linear and Nonlinear Data Dimensionality Reduction*. [Online]. <http://people.csail.mit.edu/gering/>
- [37] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 6, pp. 559-572, 1901.
- [38] T. Cox and M. Cox, *Multidimensional Scaling*, Second Edition ed. New York: Chapman & Hall, 2001.
- [39] S. T. Roweis and Saul L. K., "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, pp. 2323-2326, Dec. 2000.
- [40] J. Tenenbaum, V. de Silva, and J. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol. 290, pp. 2319-2323, Dec. 2000.
- [41] F. S. Tsai, "Dimensionality reduction techniques for blog visualization," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2766–2773, 2011.
- [42] N. Kwak, C. Kim, and H. Kim, "Dimensionality reduction based on ICA for regression problems," *Neurocomputing*, vol. 71, no. 13-15, pp. 2596-2603, August 2008.
- [43] I. Jolliffe, *Principal Component Analysis*, Second edition ed. New York: Springer-Verlag, 1986.
- [44] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent component analysis*. Canada: John Wiley & Sons, 2001.
- [45] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed.: Academic Press, 1990.
- [46] A. H. Sahoozadeh, B. Z. Heidari, and C. H. Dehghani, "A New Face Recognition Method using PCA, LDA and Neural Network," *International Journal of Electrical and Electronics Engineering*, vol. 2, no. 8, pp. 507-512, 2008.
- [47] L. H. Chan, S. H. Salleh, and C. M. Ting, "PCA, LDA and neural network for face identification," in *Industrial Electronics and Applications, ICIEA 2009. 4th IEEE Conference on*, xi'an, May. 2009., pp. 1256-1259, 25-27.
- [48] M. Soo Park, J. H. Na, and J. Y. Cho, "PCA-based Feature Extraction using Class Information," in *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, vol. 1, Oct. 2005., pp. 341- 345.
- [49] W. Jiao and Y. Chang, "Fault diagnosis of rotor systems Using ICA Based Feature Extraction ," in *Proceedings of the 2009 IEEE International Conference on Robotics and Biomimetics*, Dec. 2009. , pp. 1286-1291.
- [50] W. Liao and J. Jiang, "Image Feature Extraction based on Kernel ICA," in *Image and Signal Processing, 2008. CISP '08. Congress on*, vol. 2, Sanya, May. 2008., pp. 763-767.
- [51] F.R. Bach and M.I. Jordan, "Kernel Independent Component Analysis," *Journal of Machine Learning Research*, vol. 3, pp. 1-48., 2002.
- [52] W. Yang, L. Zhang, and C. Sun, "Fuzzy 2DLDA for Face Recognition," in *Pattern Recognition, 2009. CCPR 2009, Chinese Conference on*, Nov. 2009., pp. 1-4, 4-6.

- [53] J.H. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 165-175., Mar. 1989.
- [54] J. Yang, D. Zhang, A. F. Frangi, and J. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, vol. 26, no. 1, pp. 131-137, Jan. 2004.
- [55] H. Kong et al., "Generalized 2D principal component analysis," in *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, vol. 1, Jul. 2005., pp. 108-113.
- [56] L. Wang, X. Wang, X. Zhang, and J. Feng, "The equivalence of two-dimensional PCA to line-based PCA," *Pattern Recognition Letters*, vol. 26, no. 1, pp. 57-60, January 2005.
- [57] J. Yang and L. Chengjun, "Horizontal and Vertical 2DPCA-Based Discriminant Analysis for Face Verification on a Large-Scale Database," *Information Forensics and Security*, *IEEE Transactions on*, vol. 2, no. 4, pp. 781-792, Dec. 2007.
- [58] W.S. Zheng, J.H. Lai, and Z. Li Stan, "1D-LDA vs. 2D-LDA: when is vector-based linear discriminant analysis better than matrix-based," *Pattern Recognition*, vol. 41, no. 7, pp. 2156-2172, 2008.
- [59] W. Yan, X. Yan, L. Zhang, and C. Sun, "Feature extraction based on fuzzy 2DLDA," *Neurocomputing*, vol. 73, no. 10-12, pp. 1556-1561, Jun. 2010.
- [60] C. Deisy, B. Subbulakshmi, S. Baskar, and N. Ramaraj, "Efficient Dimensionality Reduction Approaches for Feature Selection," in *International Conference on Computational Intelligence and Multimedia Applications*, vol. 2, Sivakasi, Tamil Nadu, Dec. 2007, pp. 121-127.
- [61] M. Dash and H. Liu, "Feature Selection for Classification," *Intelligent Data Analysis: An Internat. J.*, vol. 1, no. 1, pp. 131-156, 1997.
- [62] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artificial Intelligence*, vol. 151, no. 1-2, pp. 155-176, Dec. 2003.
- [63] K. Jaber, R. Abdullah, and N.A. Rashid, "A framework for decision tree-based method to index data from large protein sequence databases," in *Biomedical Engineering and Sciences (IECBES), 2010 IEEE EMBS Conference on*, Kuala Lumpur, Nov.2011, pp. 120-125.
- [64] I.K Sethi, "Entropy nets: from decision trees to neural networks," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1605-1613, Oct. 1990.
- [65] T. Kohonen, *Self-Organizing Maps*, Third Edition ed. Berlin: Springer, 1995.
- [66] R. Kohav and G.H. John, "Wrappers for Feature Subset Selection," *Artificial Intelligence*, vol. 97, pp. 273-324, May. 1997.
- [67] M. Blachnik, W. Duch, A. Kachel, and J. Biesiada, "Feature Selection for Supervised Classification: A Kolmogorov-Smirnov Class Correlation-Based Filter," in *AI-Meth Series 2009, Symposium On Method of Artificial Intelligence*, Gliwice, Nov.2009, pp. 18-19.
- [68] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*, Prmera edición ed.: Englrwood Cliffs: Prentice Hall, 1982.
- [69] D. W. Aha and R. L. Bankert, "A Comparative Evaluation of Sequential Feature Selection Algorithms," in *Learning from data: artificial intelligence and statistics V*. New York: Springer Verlag, 1996., ch. 19, pp. 199-206.
- [70] P. Pudil, J. Novovicova, and J. Kittler, "Floating Search Methods in Feature Selection," *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1,119-1,125, Nov. 1994.
- [71] M. Kudo and J. Sklansky, "Comparison of Algorithms that Selects Features for Pattern Classifiers," *Pattern Recognition*, vol. 33, no. 1, pp. 25-41, Jan. 2000.
- [72] S. K. Pal, R. K. De, and J. Basak, "Unsupervised Feature Evaluation: A Neuro-Fuzzy Approach," *IEEE Trans. Neural Network*, vol. 11, no. 2, pp. 366-376, Mar. 2000.
- [73] K. Kira and L. Rendell, "A practical Approach to Feature Selection," in *Proc. Ninth Int'l. Workshop Machine Learning*, San Francisco, 1992., pp. 249-256.
- [74] D. Koller and M. Sahami, "Towards Optimal Feature Selection," in *Proc. 13th Int'l. Conf. Machine Learning*, 1996, pp. 284-292.
- [75] P. Mitra, P. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 24, no. 6, pp. 301-312, Jun. 2002.
- [76] L. A. Zadeh, "FuzzySets," *Information and Control*, vol. 8, pp. 338-353, 1965.
- [77] Z. Wang, Q. Wang, M Bai, Z. Chen, and Z. Sun, "Further exploration on relationship between crisp sets and fuzzy sets," in *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on*, vol. 2, 2010, pp. 609-613.
- [78] W. Pedricz, A. Skowron, and V. Kreinovich, *Handbook of Granular Computing*. England: A Jhon Wiley & Sons, 2008.
- [79] L. Zehua, J. Hai, and Y. Pingpeng, "The Theory of Triangle Type-2 Fuzzy Sets," in *Computer and Information Technology, 2009. CIT '09. Ninth IEEE International Conference on*, Oct. 2009, pp. 57-62, 11-14.

- [80] J.M. Mendel and Wu Hongwei, "Type-2 Fuzzistics for Symmetric Interval Type-2 Fuzzy Sets: Part 1, Forward Problems," *Fuzzy Systems, IEEE Transactions on*, vol. 14, no. 6, pp. 781-792, Dec. 2006.
- [81] J.M. Mendel, "Type-2 fuzzy sets and systems: An overview.," *IEEE Comput. Intell. Mag.*, vol. 2, no. 1, pp. 20-29., Feb. 2007.
- [82] H. Hagras, "Type-2 FLCs: A New Generation of Fuzzy Controllers ," *Computational Intelligence Magazine, IEEE*, vol. 2, no. 1, pp. 30-43, Feb. 2007.
- [83] J.M. Mendel, R.I. John, and Liu Feilong, "On using type-1 fuzzy set mathematics to derive interval type-2 fuzzy logic systems," in *Fuzzy Information Processing Society, 2005. NAFIPS 2005. Annual Meeting of the North American*, 2005, pp. 528- 533.
- [84] K. W. Reed, "Type-1 And Type-2 Fuzzy Systems For Detecting Visitors In An Uncertain Environment," *Dissertations, Academic, University of Missouri, Columbia*, Computer engineering, Columbia, Tesis Maestria 2009.
- [85] H. Wu, H. Wu, and J. Luo, "An Interval Type-2 Fuzzy Rough Set Model for Attribute Reduction," *IEEE Transactions on Fuzzy System*, vol. 17, no. 2, pp. 301-315, 2009.
- [86] E. C. Tsang, D. Chen, D. S. Yeung, X. Z. Wang, and J. W. Lee, "Attributes Reduction Using Fuzzy Rough Sets," *IEEE Transactions On Fuzzy Systems*, vol. 16, no. 5, pp. 1130-1141, 2008.
- [87] R. Jensen and Q. Shen, "Fuzzy-Rough Sets for descriptive Dimensionality Reduction," in *Fuzzy Systems, IEEE International Conference, Honolulu, HI , USA , 2002*, pp. 29-34.
- [88] R. B. Bhatt and M. Gopa, "On fuzzy-rough sets approach to feature selection," *Pattern Recognition Letters*, vol. 26, no. 7, pp. 965-975, May 2005.
- [89] G. C. Tsang, C. Degang, E. C. Tsang, J. W. Lee, and D. S. Yeung, "On attributes reduction with fuzzy rough sets," in *IEEE Conferences*, vol. 3, 2005, pp. 2775-2780.
- [90] P. Maji and S. K. Pal, "Feature Selection Using F-information Measures in Fuzzy Approximation Spaces," *IEEE Transaction on Knowledge and data Engineering*, vol. 22, no. 6, pp. 854-867, Jun. 2010.