Inclusión de un Índice de Estimación de Incertidumbre, Distribución y Cohesión de Datos en el Modelamiento Borroso

Including an Index for Estimating Uncertainty, Distribution and Cohesion of Data in Fuzzy Modeling

Carlos M. Sierra D.[1], MSc. y Hernán D. Álvarez Z.[2], PhD.
1. Grupo de investigación en Simulación de Comportamiento de Sistemas., Departamento de Ingeniería de Sistemas, Universidad de Antioquia,
2. Grupo de investigación en Automática. Universidad Nacional de Colombia, GAUNAL. Escuela de Procesos y Energía.
camasi064@udea.edu.co, hdalvare@unal.edu.co

Recibido para revisión 26 de Marzo de 2007, aceptado 15 de Junio de 2007, versión final 26 de junio de 2007

Resumen— Este trabajo presenta una propuesta de estimación de la incertidumbre y la calidad de agrupamiento producidos en la identificación de modelos mediante Sistemas de Inferencia Borrosa del tipo Takagi-Sugeno (SIB T-S). Además, propone la integración de tales medidas como criterios evaluadores del modelo con base en la incertidumbre y la partición borrosa generados durante su obtención. Tal estimación hará que el modelo obtenido sea el de menor incremento en la incertidumbre frente a los datos originales del proceso. Además, permite evaluar la distribución y densidad de los datos en los conjuntos borrosos obtenidos durante el modelamiento usando SIB T-S. Los valores de tal índice pueden ser usados como complemento al modelo final cuando este es usado en cualquier tarea basada en modelo (diseño, optimización, control, etc.). Esas tareas suponen un modelo con incertidumbre uniforme del modelo (que se asume baja), en todo el espacio del modelo. Usando el índice propuesto, se puede calcular un valor más realista de la incertidumbre del modelo en cualquier punto del espacio del modelo.

Palabras Clave— Controladores Basados en Modelo, Identificación de Sistemas, Manejo de Incertidumbre, Modelamiento de Información Y Técnicas De Razonamiento, Sistema de Inferencia Borrosa Takagi-Sugeno.

Abstract— This paper presents a proposal for estimating the uncertainty and grouping quality that take place when a model is identified using a Takagi-Sugeno Fuzzy Inference System (T-S FIS). Additionally, the integration of such measures as criteria for model evaluation based on uncertainty and fuzzy partition generated during model identification is proposed. Such an index allows to identifying a model that causes a minimum uncertainty increment respecting original process data. Additionally, the index evaluates data distribution and density at obtained fuzzy sets during fuzzy modeling. The index values can be used as a complement to the final model when it is used in any model-based task (design, optimization, control, etc). Such class of tasks supposes a model with uniform uncertainty (assumed low) in all model space. Using proposed index a more realistic model uncertainty value may be calculated at any point in the model space.

Key words— Model-Based Controller Design, System Identification, Takagi-Sugeno Fuzzy Inference Systems, Uncertainty Management, Uncertainty Modeling and Reasoning Techniques.

I. INTRODUCCION

La inclusión de la incertidumbre en la modelación de fenómenos cada vez toma mayor preponderancia. De los modelos determinísticos, representaciones con componentes e interrelaciones definidas y completamente predecibles, se ha pasado a los estadísticos, convenientes para modelar fenómenos en los que la interacción no puede predecirse exactamente. Es conveniente, sin embargo, resaltar el carácter poco realista de un modelo totalmente determinístico, puesto que un fenómeno modelado puede verse afectado por acciones de naturaleza aleatoria provenientes de su entorno.

Además, puede existir un comportamiento no determinístico en la intensidad de las interacciones de sus componentes. Esto explica porque gran cantidad de fenómenos no pueden considerarse ni determinísticos ni completamente aleatorios. Para estos fenómenos se han propuesto en las últimas décadas, modelos que intentan representar su incertidumbre y, a la vez, manejar un grado de complejidad aceptable para un usuario humano. Ahora bien, en ingeniería es ineludible la obtención de un modelo matemático para aproximar un sistema real. Lo ideal sería que este modelo se obtuviese de acuerdo con los principios físicos que lo rigen. Sin embargo, esto no siempre es posible y debe recurrirse a un proceso conocido como identificación de sistemas, mediante el cual se obtiene un modelo matemático del sistema a partir de un conjunto de muestras entrada-salida [1]. Una opción interesante es la deducción de un modelo basado en el principio de conservación (masa, energía y cantidad de movimiento), así sea aproximado, que se emplee como generador de datos para hallar un modelo mucho más operativo desde el punto de vista computacional.

Tras una revisión de las medidas de incertidumbre reportadas en la literatura, este trabajo presenta una propuesta de estimación de la incertidumbre introducida en la identificación modelos mediante Sistemas de Inferencia Borrosa tipo Takagi-Sugeno (SIB T-S). Además, propone la integración de tal medida en un proceso de identificación de un modelo usando SIB T-S, como un criterio evaluador del modelo con base en la incertidumbre generada durante su obtención. Tal estimación hará que el modelo obtenido sea el de menor incremento en la incertidumbre frente a los datos originales. Además, la medida será un valor agregado para el modelo final, a la hora de ser usado en tareas de control, en las que el acotamiento de tal incertidumbre es crucial en el diseño de controladores basados en modelo.

Igualmente importante será tomar en consideración qué tan adecuada ha sido la partición borrosa que se efectúa en un proceso de identificación. Cuando esta partición es obtenida mediante un agrupamiento borroso de datos, características tales como la repartición adecuada de los datos en los diferentes Conjuntos Borrosos Multi-Dimensionales (CBMDs) y la cohesión de estos CBMDs como una medida de la similaridad de los datos según su pertenencia a ellos, se vuelven importantes para evaluar el resultado de la identificación. Aunque este artículo se despliega un poco más en el tema de la incertidumbre, el de la evaluación de la calidad del agrupamiento es sensiblemente importante como propuesta que ha sido elaborada al interior de la tesis doctoral *Incertidumbre en el Proceso de Identificación de Sistemas de Inferencia Borrosa Takagi-Sugeno*, actualmente en desarrollo.

II. INCERTIDUMBRE EN UN PROCESO DE IDENTIFICACIÓN

La concurrencia de los temas de modelación y de manejo de incertidumbre se hace evidente en la ingeniería. La identificación de sistemas, como medio para obtener un modelo matemático de un sistema, es un tema frecuentemente aludido en el análisis de procesos o sistemas físicos. La identificación se pone en práctica cuando no es posible o es demasiado complejo obtener un modelo fenomenológico, o de *primeros principios*, de un sistema real. La identificación tiene como fin, luego de seleccionada una estructura apropiada (expresión matemática general), determinar unos valores adecuados de los parámetros de la misma con base en la medición de datos de entrada-salida del sistema. Aunque la aplicación de técnicas de identificación determinísticas es bastante frecuente, también se están abriendo paso las técnicas que toman en cuenta la incertidumbre presente en los fenómenos y sus modelos. De esta manera se busca que los modelos intermedios, y final, del proceso, evalúen su capacidad de representación a luz de la cantidad de información pérdida (lo que es igual, a la cuantía de incertidumbre agregada) y no sólo con base en la diferencia entre las salidas determinadas por el modelo y el sistema. Toda expresión matemática elegida para representar un sistema en un proceso de identificación, es una simplificación del mismo, y por ello pierde parte de la información que posee completa el sistema. La aproximación conseguida con un modelo tiene como fin obtener niveles manejables de las complejidades *descriptiva* y de *incertidumbre* [2]. Estos tipos de complejidad, pueden ser utilizados como criterios de comparación y elección en el problema de *la simplificación de sistemas*.

En el proceso de obtener un modelo siempre habrá pérdida de la información asociada con cada tipo de complejidad. La relación de estos dos tipos de complejidad, en el contexto de la simplificación de sistemas, ha sido establecida en [3] así: "Cuando simplificamos un sistema, queremos reducir tanto la complejidad basada en la información descriptiva como la basada en la información de incertidumbre. Desafortunadamente, estas dos complejidades entran en conflicto una con otra. En general, cuando reducimos una, la otra se incrementa o, al menos, permanece inmodificable". En términos llanos, la *complejidad descriptiva* se relaciona con la cantidad de objetos y la variedad de interacciones que estipula el modelo. En los modelos matemáticos los objetos se hacen corpóreos a través de variables (que representan algunos de sus atributos), y las interacciones aparecen como operaciones o funciones que enlazan esas variables. La *complejidad de incertidumbre* aparece cuando debe elegirse una(s) de entre un conjunto de alternativas posibles, o cuando esas alternativas entran en conflicto, como en el caso de los modelos estocásticos. A cada tipo de complejidad se le asocia su correspondiente tipo de información. La *información descriptiva* es aquella que utiliza el modelo para representar el sistema. La *información de incertidumbre* es aquella que debe recogerse para resolver cualquier incertidumbre presentada en el modelo. La bondad o calidad de un modelo, sea intermedio o final, puede evaluarse considerando los grados de su complejidad descriptiva (calculada en función de su información descriptiva) y de su complejidad de incertidumbre (estimada según la cantidad de información de incertidumbre).

Debido a la importancia que ha ido adquiriendo la incertidumbre en el ámbito de la modelación, este artículo profundiza sólo en el tema de la complejidad de incertidumbre. A continuación se detallan las fuentes que originan incertidumbre en un proceso de identificación.

El uso de datos incorrectos o incompletos. Los datos empleados en un proceso de identificación de un sistema, pueden ser imprecisos, incorrectos o incompletos. Las dos primeras deficiencias pueden deberse a su obtención mediante procedimientos erróneos, o a través de instrumentos deficientes (defectuosos, de baja precisión, o mal calibrados), o debidos a la intrusión de ruido en los valores de las señales medidas. La incompletés de los datos puede ser causada por un inapropiado diseño de experimentos, en el que o no se pudieron identificar las variables de control necesarias (introduciendo perturbación en los valores de las señales muestreadas) o en el que no se consideraron las condiciones experimentales suficientes para descubrir sus dinámicas relevantes (principales y secundarias). Estas alteraciones introducen incertidumbre acerca de los valores correctos y de la cantidad necesaria de los datos entrada-salida para el proceso de identificación, lo que evidencia su conexión con la complejidad de incertidumbre introducida en el modelo.

La utilización de un método de identificación. Los métodos de identificación se alimentan con datos de entrada-salida del sistema, junto con la incertidumbre presente en ellos, la cual puede propagarse e incrementarse a través de las distintas etapas del método. La imperfección de los datos afecta fundamentalmente los valores de los parámetros funcionales de la estructura (en algunos tipos de modelo también influye en los valores de los parámetros estructurales). Por otro lado, un método de identificación entendido como una sucesión de etapas definidas rigurosamente en términos de expresiones matemáticas, pretende concretar la estructura o expresión matemática elegida que aproximará al sistema. Esta aproximación deriva en una reducción de la complejidad descriptiva. Sin embargo, tal reducción de complejidad descriptiva repercute en un aumento de incertidumbre en el modelo obtenido.

Lo trascendente en el tema de la simplificación de sistemas, cuando ella se realiza a través del criterio de la complejidad de incertidumbre, es responder la pregunta: ¿cómo puede representarse y calcularse la incertidumbre?. Este será el tema desarrollado en los siguientes dos apartados.

III. TEORÍAS DE INCERTIDUMBRE

Dos grandes clases de incertidumbre se han formulado hasta el momento, la *incertidumbre debida a la deficiencia de información*, conocida también como **ambigüedad**, y la *incertidumbre lingüística* o **vaguedad**. La primera de ellas ha sido desarrollada al interior de la teoría de los conjuntos clásicos. La segunda ha evolucionado en el ámbito de la teoría de los conjuntos borrosos [4]. En términos muy generales, la ambigüedad se presenta cuando con información deficiente **debe elegirse una de entre una colección** de alternativas claramente definidas y mutuamente excluyentes (la exclusión en este contexto significa que sólo puede elegirse una de las alternativas). En la vaguedad, la incertidumbre se origina por una **deficiente especificación de cada alternativa**. Para posibilitar su tratamiento matemático, entiéndase ambigüedad como la dificultad para determinar el subconjunto al cual pertenece un elemento del cual solo se dispone de una pobre evidencia o información. En el tratamiento formal de la vaguedad, las alternativas se representan por medio de conjuntos borrosos, lo que resalta la dificultad de discriminación. La investigación sobre ambigüedad ha permitido establecer dos subclases: el **conflicto** y la **no-especificidad**. La evidencia es conflictiva en el momento en que apunta a varios subconjuntos. Si estos subconjuntos pueden ordenarse de acuerdo con la relación de inclusión, el conflicto será mínimo. (El conflicto será nulo si la evidencia indica un solo subconjunto). La no-especificidad alude al tamaño de los subconjuntos, entre más elementos contengan los subconjuntos, menos específica es la evidencia. Así pues, una evidencia perfecta (no conflictiva y específica) señalaría un único subconjunto unitario. La figura 1, muestra la taxonomía de los tipos de incertidumbre descritos.

El conflicto es el tipo de incertidumbre más investigado. La única teoría que se ha desarrollado para representar el conflicto y la más conocida de todas las que componen el ámbito de la incertidumbre en general, es la *Teoría de la probabilidad.* La teoría de la probabilidad evoluciona alrededor de un tipo de medida, aditiva por demás, la medida de probabilidad. Para la no-especificidad, en cambio, se han planteado varias teorías, todas ellas fundamentadas en las denominadas medidas no-aditivas. Es el caso de la *Teoría de la evidencia*, basada en las medidas de plausibidad y credibilidad.

Otra teoría para el tratamiento de la no-especificidad es la *Teoría de la posibilidad*, cimentada sobre la interpretación de las funciones de pertenencia como distribuciones de posibilidad. También se encuentra la *Teoría de la posibilidad como caso especial de la teoría de la evidencia*, que se establece según las medidas de posibilidad y necesidad. Por último está la evolución alrededor de las diversas *Teorías de las probabilidades imprecisas.* Las diferentes teorías de las probabilidades imprecisas se apoyan en las medidas de probabilidad superior e inferior. A continuación se hará una presentación resumida de estas teorías en conjuntos universo finitos.

Resta por mencionar que la única teoría que se ha formulado para el tratamiento formal de la vaguedad es la *Teoría de los conjuntos borrosos.*

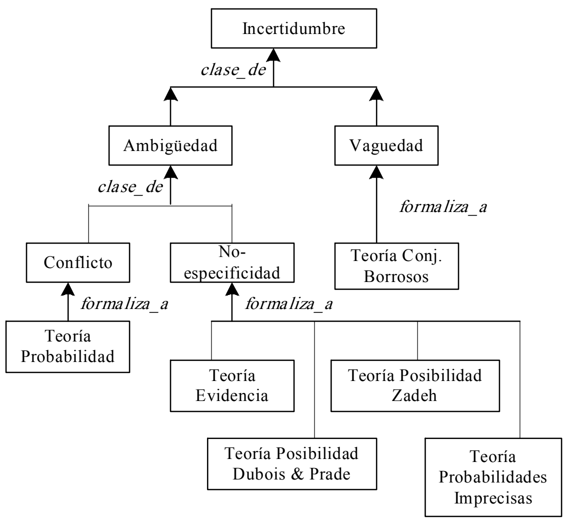

Figura 1. Representación taxonómica de la incertidumbre, junto con las teorías que las representan.

A. Teoría de la probabilidad

Esta teoría se originó con los estudios desarrollados por Pascal y Fermat en 1654, y su aplicación inicial en el contexto de los juegos de azar [5]. Dado un conjunto universo finito X, la medida de probabilidad de $A \in P(X)$, $\Pr(A)$, se estima como: $\Pr(A) = \sum_{x \in A} p(x)$, donde: $P(X)$ es el conjunto potencia de X y $\mathrm{p} = \left\langle p(x) \mid p(x) \in [0,1], \sum_{x \in X} p(x) = 1 \right\rangle$ es una distribución de probabilidad sobre X. Además: *i)* $\Pr(\phi) = 0$ y $\Pr(X) = 1$ y *ii)* $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ siempre que $A \cap B = \phi$.

B. Teoría de los conjuntos borrosos

Teoría cuyo trabajo seminal se presentó en [6]. Sea X un conjunto universo finito. En la teoría de conjuntos borrosos, la pertenencia de los elementos de X a un subconjunto borroso $A \in P(X)$, se denota a través de una función de pertenencia, μ_A, tal que: $\mu_A : X \to M$, donde el conjunto M se denomina *conjunto de pertenencia asociado*. Cuando $M = [0,1]$, el conjunto borroso A se denomina estándar. El valor $\mu_A(x)$ se le conoce como grado de pertenencia. Una forma común de simbolizar un conjunto borroso es: $A = \sum_{x \in X} \mu_A(x)/x$, donde $\sum$ se emplea para representar la unión de los subconjuntos unitarios de X. La *altura* de un conjunto borroso A se define como: $h(A) = \max_x \mu_A(x)$. Cuando $h(A) = 1$, al conjunto A se lo llama *normal*, de lo contrario *subnormal*. Al conjunto no borroso, o concreto, ${}^{\alpha}A = \{x \mid \mu_A(x) \geq \alpha\}$ se le denomina α-corte de A.

C. Teoría de la Evidencia.

Esta teoría se planteó y refinó en [7] y [8], respectivamente. Se basa en dos medidas, la plausibidad y la credibilidad. Dado un conjunto universo finito X, la medida de plausibidad de $A \in P(X)$, $\mathrm{Pl}(A)$, se calcula como: $\mathrm{Pl}(A) = \sum_{B \mid B \subseteq A} m(B)$, donde m es una función conocida como *asignación de probabilidad básica* que: *i)* $m(A) \in [0,1], \forall A \in P(X)$, *ii)* $m(\phi) = 0$ y $\sum_{B \in P(X)} m(B) = 1$, *iii)* $\mathrm{Pl}(A \cup B) \leq \mathrm{Pl}(A) + \mathrm{Pl}(B)$ siempre que $A \cap B = \phi$. Se define como *cuerpo de evidencia* a la pareja $\langle \mathcal{F}, m \rangle$, donde $\mathcal{F} = \{A \mid A \in P(X), m(A) > 0\}$, los subconjuntos $A \in \mathcal{F}$ se denominan *elementos focales*. La medida de credibilidad se calcula como: $\mathrm{Bel}(A) = 1 - \mathrm{Pl}(\overline{A})$, $\forall A \in P(X)$, donde $\overline{A}$ es el conjunto complemento de A.

D. Teoría de la posibilidad formulada por Zadeh

En la teoría de la posibilidad presentada en [9], la información es expresada mediante proposiciones borrosas. Estas proposiciones toman la forma χ es *F*, donde χ es una *variable lingüística*, y *F* es el rótulo de un subconjunto borroso normal y estándar del conjunto finito *X*. Se define una *variable borrosa V*, asociada con χ, la cual puede tomar como valor a alguno de los elementos del conjunto universo finito *X*, pero restringido de acuerdo con el conjunto borroso que representa *F*. La expresión χ es *F* se traduce en una asignación de la función de pertenencia del conjunto borroso *F*, μ_F, a la función de distribución de posibilidad de los valores que **puede** asumir la variable *V*, r_F, o sea: $\mu_F = r_F$, donde $r_F(x)$ es el *grado de posibilidad* que $V = x$, $\forall x \in X$, con base en la evidencia disponible (en este caso la evidencia disponible se expresa mediante la proposición χ es *F*).

La medida de posibilidad para un subconjunto concreto *A*, inducida por la proposición borrosa χ es *F*, $\mathrm{Pos}_F(A)$, se calcula por medio de: $\mathrm{Pos}_F(A) = \sup_{x \in A} r_F(x)$, $\forall A \in P(X)$

La medida de posibilidad en el caso que $A \in P(X)$, y A sea borroso: $\mathrm{Pos}_F(A) = \sup_{x \in X} \min\left[r_F(x), \mu_A(x)\right]$.

La medida de necesidad inducida por la proposición borrosa χ es *F*, para cualquier subconjunto A de *X*, $\mathrm{Nec}_F(A)$, puede obtenerse como: $\mathrm{Nec}_F(A) = 1 - \mathrm{Pos}_F(\overline{A})$, $\forall A \in P(X)$. Además: $\mathrm{Pos}_F(A \cup B) = \max\left[\mathrm{Pos}_F(A), \mathrm{Pos}_F(B)\right]$.

E. Teoría de la posibilidad formulada por Dubois y Prade

Esta teoría fue planteada en [10]. Como se mencionó anteriormente, se cimienta en las medidas de posibilidad y necesidad. Esta teoría puede verse como un caso de la teoría de la evidencia, en la que los *elementos focales* son *anidados*, es decir $\mathsf{F} = \{A_1, A_2, ..., A_n\}$ en el que $A_1 \subset ... \subset A_n$, con $card(\mathsf{F}) = |\mathsf{F}| = n$ el número de elementos focales. Dado un conjunto universo finito X, la medida de posibilidad, $\text{Pos}(A) \quad \forall A \in P(X)$, se estima como: $\text{Pos}(A) = \max_{x \in A} r(x)$, donde r es una *función de posibilidad*. Una manera de determinar la función r es organizando los elementos de X de modo que quede de manifiesto $A_1 \subset ... \subset A_n$, en el que $A_i = \{x_1, ..., x_i\} \quad i \in \mathrm{N}_n = \{1, 2, ..., n\}$. Con base en lo anterior, puede demostrarse que: $r(x_i) = \sum_{k=i}^{n} m(A_k)$. Además, $\text{Pos}(A \cup B) = \max[\text{Pos}(A), \text{Pos}(B)] \quad \forall A, B \in P(X)$. La *medida de necesidad*, $\text{Nec}(A) \quad \forall A \in P(X)$, se determina como: $\text{Nec}(A) = 1 - \text{Pos}(\bar{A})$.

F. *Teoría de las probabilidades imprecisas*

Existen varias de estas teorías, sin embargo, todas ellas comparten el fundamento conceptual y matemático presentado en [11]. En estas teorías la evidencia queda totalmente especificada por una *función de probabilidad inferior* $\underline{g}$ (o alternativamente por una *función de probabilidad superior* $\bar{g}$). La función $\underline{g}$ debe cumplir $\sum_{x \in X} \underline{g}(x) \leq 1$. Suponga que existe una función m_D para la que $m_D(\phi) = 0$ y $\sum_{A \in P(X)} m_D(A) = 1$ en la que D es un conjunto convexo y cerrado de funciones de probabilidad. La función de probabilidad inferior con base en este conjunto D, se expresa como: $\underline{g}_D(A) = \sum_{B \mid B \subseteq A} m_D(B), \quad \forall A \in P(X)$. La función de probabilidad superior con base en el conjunto D, se calcula así: $\bar{g}_D(A) = 1 - \underline{g}_D(\bar{A}), \quad \forall A \in P(X)$.

Además, siempre que $A \cap B = \phi$, se tiene: $\underline{g}_D(A \cup B) \geq \underline{g}_D(A) + \underline{g}_D(B) \quad \forall A, B \in P(X)$.

IV. MEDIDAS DE INCERTIDUMBRE

Una expresión matemática que pretenda ser utilizada para calcular la ambigüedad (sea el conflicto o la no-especificidad) presente en un sistema, debe cumplir los requisitos axiomáticos expuestos en [12]. No obstante, la concreción de estas medidas dependerá de la teoría de incertidumbre en la que se implante. Del mismo modo en [2], se han establecido las condiciones necesarias para que alguna función matemática pueda estimar la vaguedad. Inicialmente serán presentadas las medidas para calcular la no-especificidad y el conflicto. Se finalizará este aparatado con diferentes formulaciones para estimar la vaguedad.

A. *Medidas para estimar la no-especificidad*

Medida Hartley. Presentada en [13], fue la primera medida formulada para calcular la no-especificidad. Considérese una situación en la que se necesita identificar la alternativa verdadera, de entre un conjunto finito X de alternativas mutuamente excluyentes. Asúmase que de acuerdo con la evidencia poseída, se conoce que las alternativas posibles se encuentran en un subconjunto $E \subseteq X$, y aquellas imposibles, se hallan por fuera de E. Lo anterior se puede formalizar mediante una distribución de posibilidad $r_E = \langle r_E(x) \mid x \in X \rangle$, $r_E(x) = \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases} \quad \forall x \in X$.

La medida de no-especificidad de acuerdo con la función Hartley, se calcula como: $H(r_E) = c \log_b |E|$. Lo común es emplear $c = 1$ y $b = 2$, de manera que la unidad de medición es el bit.

Medida U-incertidumbre. Esta medida fue planteada en [14] con el fin de generalizar la medida Hartley. En realidad, propusieron varias de estas medidas de acuerdo con la teoría de incertidumbre en la que se ubique. Una de las formas de la medida *U*-incertidumbre la plantearon para la visión de posibilidad hecha por Zadeh. En este contexto, la *U*-incertidumbre asume la forma: $U(r_F) = \int_0^1 \log_2 |{}^{\alpha}F| d\alpha$, donde $|{}^{\alpha}F|$ es la cardinalidad del conjunto α-corte, ${}^{\alpha}F$.

Para el enfoque de la teoría de la posibilidad dado por Dubois y Prade, la medida de *U*-incertidumbre se concreta como: $U(r) = \sum_{i=2}^{n} (r(x_i) - r(x_{i+1})) \log_2 i$ donde $r = \langle r_1, ..., r_n \rangle$ es una distribución de posibilidad ordenada tal que $1 = r_1 \geq r_2 \geq ... \geq r_n$, ($r_{n+1} = 0$ por convención).

La versión de la medida *U-incertidumbre* para ser aplicada en la Teoría de la evidencia, fue planteada en [15]. En esta teoría la estimación de la no-especificidad se obtiene mediante:

$$U(m) = \sum_{A \in F} m(A) \log_2 |A|.$$

Medida N-incertidumbre. Este avance realizado al interior de las diversas teorías de las probabilidades imprecisas fue expuesto en [16]. La *N-incertidumbre*, se calcula a través de $N(C) = \sum_{A \in P(X)} m(A) \log_2 |A|$, en la que la evidencia disponible se formula mediante un conjunto convexo de funciones de distribución de probabilidad, *C*.

$$C=\left\langle p \mid p(x)\in[0,1] \quad \forall x\in X, \sum_{x\in X} p(x)=1, \quad \mathrm{Bel}(A)\leq \sum_{x\in A} p(x)\leq 1-\mathrm{Bel}(A)\right\rangle$$

El cálculo de $N(C)$ requiere obtener la asignación de probabilidad básica de cada subconjunto de X $m(A)=\sum_{B|B\subseteq A}(-1)^{|A-B|}f(B) \quad ,\forall A\in P(X)$, en la que $f(A)=\inf_{p\in C}\Pr(A), \ \forall A\in P(X)$.

B. Medidas para estimar el conflicto

Entropía de Shannon. Esta medida fue formula en [17] y es la única medida que puede estimar de manera directa la cantidad de conflicto en un sistema. Esta medida se inserta dentro de la teoría de la probabilidad. La expresión para calcular la entropía Shannon es: $S(p)=-\sum_{x\in X}p(x)\log_b p(x)$, en la que $b=2$ para que la unidad sea el bit y $p=\langle p(x)\mid x\in X\rangle$ es una distribución de probabilidad sobre X. Se demostró que a diferencia de la no-especificidad, el conflicto no posee una medida que la calcule de manera directa en ninguna otra teoría distinta de la teoría de la probabilidad. Sin embargo, la estimación del conflicto puede llevarse a cabo de manera indirecta. La propuesta es que primero se calcula la *incertidumbre agregada*, AU (el conflicto y no-especificidad como un todo) de la cual se sustrae el valor de la no-especificidad, para obtener el valor del conflicto. En la actualidad, el único funcional que ha pasado las pruebas para calcular la incertidumbre agregada, es: $AU(\mathrm{Bel})=\max_{D_{\mathrm{Bel}}}\left\{-\sum_{x\in X}p(x)\log_2 p(x)\right\}$

$$D_{\mathrm{Bel}}=\left\langle p \mid p(x)\in[0,1] \quad \forall x\in X, \sum_{x\in X} p(x)=1, \quad \mathrm{Bel}(A)\leq \sum_{x\in A} p(x)\leq 1-\mathrm{Bel}(A)\right\rangle$$

Así entonces, una medida generalizada para el conflicto se obtiene como: $GS=AU-GH$. Donde GS representa la estimación de incertidumbre para el conflicto, y GH es una medida generalizada para estimar la no-especificidad, o sea: GH es cualquiera de las medidas: $U(r_F)$, $U(r)$, $U(m)$ o $N(C)$.

C. Medidas para estimar la vaguedad

En general las medidas de vaguedad se dividen en tres clases. En un primer grupo las medidas de vaguedad dependen de la función de pertenencia del conjunto borroso y de la de su conjunto concreto más cercano. Un segundo grupo considera la vaguedad de un conjunto borroso en términos de su función de pertenencia y de la de su conjunto borroso complemento, y el tercero, sólo cuenta con el valor de pertenencia al conjunto borroso para calcular su vaguedad.

Vaguedad de un conjunto borroso en términos de su función de pertenencia y de la de su conjunto concreto más cercano.

Un conjunto de medidas de borrosidad, conocidas como *índices de borrosidad* de A, $f_w(A)$, se calculan con base en la distancia Minkowski entre el conjunto borroso A y su conjunto concreto más cercano CA, así: $f_W(A)=D_W(A,CA)$; donde CA es el conjunto concreto más cercano a un conjunto borroso A, definido mediante su *función característica* como:

$$\mu_{CA}(x)=\begin{cases}0 & \text{si} \quad \mu_A(x)<0{,}5\\ 1 & \text{si} \quad \mu_A(x)\geq 0{,}5\end{cases}$$

$\mu_{CA}(x)$, es la pertenencia del elemento x en el conjunto concreto CA. D_w, es la fórmula de Minkowski para hallar la distancia entre el subconjunto borroso A y el conjunto concreto CA. La distancia Minkowski entre dos conjuntos borrosos A y B cualquiera del conjunto universo X se calcula como:

$$D_w(A,B)=\left[\sum_{x\in X}\left|\mu_A(x)-\mu_B(x)\right|^{w}\right]^{1/w}$$

Donde $w\in[1,\infty)$ y $\mu_A(x)$ es el grado de pertenencia del elemento x en el conjunto borroso A. Existen, entonces, tantas distancias como valores posibles de w, es el caso de los siguientes.

Índice de borrosidad lineal. Este índice es una versión normalizada de $f_1(A)$: $\hat{v}(A)=\dfrac{2}{|X|}D_1(A,CA)$

Índice de borrosidad cuadrático. Similar al anterior, pero es la versión normalizada de $f_2(A)$:

$$\hat{\eta}(A)=\frac{2}{|X|^{1/2}}D_2(A,CA)$$

Vaguedad de un conjunto borroso en términos de su función de pertenencia y de la de su conjunto complemento.

Un primer conjunto de este tipo de medidas de borrosidad fue concebido en [18]:

$f_{LT}(A)=c\sum_{x\in X}\mu_A(x)\log_b\mu_A(x)+\mu_{\overline{A}}(x)\log_b\mu_{\overline{A}}(x)$. Es frecuente que $c=-1$ y $b=2$.

En la Ref. [19] se indica medir borrosidad, de la siguiente forma:

$$f_{Y,D}(A)=D(Z,\overline{Z})-D(A,\overline{A})$$

En la que D es una distancia métrica general, Z denota cualquier subconjunto concreto arbitrario del conjunto universo X, de manera que $D(Z,\overline{Z})$ es la distancia más grande posible estimada con la operación complemento utilizada.

Usando la clase de distancia Minkowski, entonces se podrían usar las siguientes expresiones:

$$D_W(A,\bar{A}) = \left[\sum_{x \in X} \left| \mu_A(x) - \mu_{\bar{A}}(x) \right|^W \right]^{1/W}, \text{ y } D_W(Z,\bar{Z}) = |X|^{1/W}.$$

Vaguedad de un conjunto borroso en términos de su función de pertenencia únicamente.

Una medida de borrosidad que sólo depende del conjunto borroso mismo, se propuso en [20]. El autor plantea evaluar la borrosidad, en un conjunto universo finito de $|X|$ elementos, como:

$$K(A) = -\sum_{x \in X} \gamma_A(x) \ln \gamma_A(x), \text{ y } \gamma_A(x) = \mu_A(x) \Big/ \sum_{x \in X} \mu_A(x)$$

V. INCERTIDUMBRE EN UN PROCESO DE IDENTIFICACIÓN DE UN MODELO USANDO SISTEMAS DE INFERENCIA BORROSA TIPO TAKAGI-SUGENO (SIB T-S)

Un SIB se emplea frecuentemente como estructura en un proceso de identificación de un sistema. En general, un SIB se compone de un conjunto de expresiones similares a las reglas lógicas Si-Entonces y de un procedimiento, o inferencia, que dispone de esas reglas para procurar un resultado cuando se le ingresan un conjunto de datos de entradas. El antecedente, y también el consecuente, de cada regla se componen de una combinación lógica de proposiciones borrosas. Cada proposición se representa matemáticamente por medio de un conjunto borroso.

En el tipo de modelo SIB T-S, formulado en [21], el antecedente de las reglas pretende aproximar las restricciones existentes entre los valores de las variables de entrada mediante una serie de conjuntos o grupos borrosos. El consecuente intenta establecer las relaciones no borrosas (concretas) entre los valores del facto de pertenencia del dato a los conjuntos de entrada y la salida correspondiente del modelo.

El trabajo doctoral *Incertidumbre en el Proceso de Identificación de Sistemas de Inferencia Borrosa Takagi-Sugeno* propone considerar la complejidad de incertidumbre generada en un proceso de modelado e identificación de SIB T-S, con el fin de obtener no sólo modelos con un buen grado de aproximación de respuesta, sino también con un nivel alto de conservación de información (información suministrada por los datos muestreados para el proceso de modelado).

El trabajo descrito a través de este artículo, no pretende abarcar todas las etapas del proceso de identificación en los que es posible perder información (agregar incertidumbre). Estas etapas se muestran en la Fig. 2. y detallan en [22]. Es el caso de la pérdida de información a través del proceso de muestreo de datos. Aunque en un proceso de muestreo los datos espurios son inevitables, se debe intentar reducir la influencia de estos en la obtención de un modelo. En este trabajo, tal influencia será disminuida a través de una suposición trascendental: *los datos empleados para la obtención del modelo ya tiene la máxima cantidad de información posible, o lo que es igual, se les ha removido la mayor cantidad factible de incertidumbre.* Esta premisa puede apoyarse en alguna(s) de las siguientes consideraciones: *i)* los datos entrada-salida han sido sometidos a un proceso de filtrado previo que reduce la influencia del ruido, *ii)* los datos han sido adquiridos ya sea a través de un proceso de simulación en un modelo fenomenológico preexistente, a los que se calificarán como *datos sintéticos* y *iii)* los datos son el resultado de someter al sistema a *señales persistentemente excitantes*. Estas apreciaciones implicarán la reducción de la complejidad de incertidumbre agregada en un modelo, debida a los datos entrada-salida. En términos concretos, los datos entrada-salida no exhibirán conflicto, no-especificidad ni vaguedad. Si ΔU_d simboliza la incertidumbre introducida por cualquier fenómeno que altere la veracidad o completés de los datos, se presumirá reducida esta alteración hasta un valor suficientemente pequeño como para desechar su influencia, de este modo $\Delta U_d = 0$.

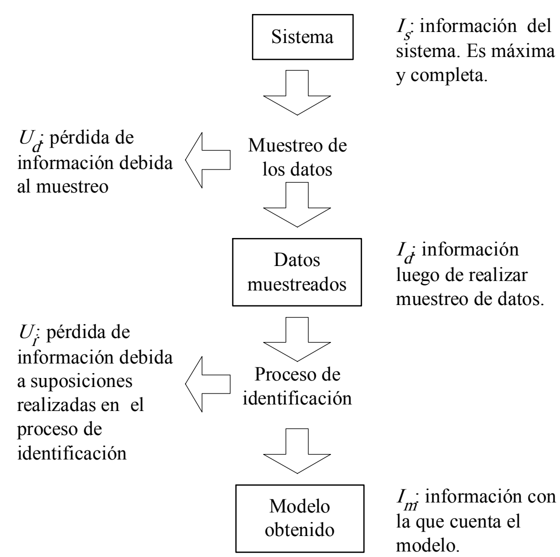

Figura 2. Flujo de información en el modelado de sistemas

El efecto del método de identificación en la obtención de un SIB T-S, se vería reflejado específicamente en la incertidumbre sobre *i)* la posición y forma del conjunto borroso del antecedente de cada regla (parámetros funcionales), *ii)* el valor inicial y la pendiente de cada función concreta de los consecuentes (parámetros funcionales) y *iii)* el número de reglas del modelo (parámetro estructural). Sea ΔU_i La cantidad de incertidumbre añadida por el método de identificación, en la que sólo se considera la incertidumbre plasmada en *i)*. De acuerdo con ello: $I_m = I_s - \Delta U_d - \Delta U_i$ y dado que se presumió $\Delta U_d = 0$, entonces: $I_m = I_s - \Delta U_i$. El propósito final del proceso de identificación es hacer que ΔU_i.sea lo más pequeño posible (idealmente: $\Delta U_i \rightarrow 0$).

El trabajo doctoral se enfoca en la modificación de la propuesta de identificación descrita en la serie de artículos [23]-[25], de modo que se inserte en él un indicador de la incertidumbre agregada en cada ciclo del proceso. Este índice servirá para evaluar y seleccionar de entre los modelos intermedios, aquel que será el modelo final del proceso de identificación.

A. Incertidumbre agregada en el método de identificación propuesto por Peña, Álvarez y otros

En el método de identificación referenciado, pueden descubrirse por lo menos dos etapas, la segunda dependiente de la primera, en las que es posible introducir incertidumbre (perder información), ellas son: el agrupamiento de reglas básicas en Conjuntos Borrosos Multi-Dimensionales (CBMD) y la proyección de estos para la determinación de los antecedentes de cada regla.

¿Por qué se presenta esta pérdida de información?. Previo a la etapa de agrupamiento, los datos de muestreo tienen la información máxima y completa del sistema. Cada uno de esos datos se interpreta como una *regla básica*. Un modelo visto así es bastante complejo descriptivamente hablando (pues hay una regla por cada dato), pero no exhibe incertidumbre: *cuando se alimenta al modelo con una entrada, sólo se activa una de todas las reglas, ofreciendo una única respuesta*. (Se omite la discusión acerca de la correctitud de la respuesta, pues eso lo evalúa el índice de desempeño). Ahora, durante la etapa de agrupamiento cada modelo generado (intermedio o final) reduce el número de reglas requeridas (a una por cada grupo, o CBMD, encontrado). Aunque la complejidad descriptiva con respecto al modelo de reglas básicas disminuye, la incertidumbre se incrementa: *cada regla participa con una posible respuesta. Habrá entonces tantas respuestas como reglas*. Esta incertidumbre sobre la salida del modelo se resuelva a través de un proceso de concreción.

Cada CBMD determinado en el proceso de agrupamiento modela un régimen de operación del fenómeno real, luego es evidente la importancia de detectar estas agrupaciones lo mejor posible. Como es sabido, la teoría de los conjuntos borrosos es utilizada para la representación de la vaguedad, y con la interpretación de la función de pertenencia como una función de posibilidad, $r_F = \mu_F$, la conexión con la no-especificidad se ha hecho clara. De acuerdo con lo anterior, en un conjunto borroso coexisten la vaguedad y la no-especificidad como tipos de incertidumbre tangibles. La inquietud que surge en este punto es: ¿también puede manifestarse el otro tipo de incertidumbre faltante...el conflicto?.

De acuerdo con la fórmula $GS = AU - GH$, se puede obtener un estimativo del conflicto, *GS*. Luego en cada conjunto borroso podría estimarse el conflicto y, en consecuencia, determinar los tres tipo de incertidumbre: la vaguedad, el conflicto y la no-especificidad. Teóricamente, al interpretar un conjunto borroso como la unión de sus α-cortes (los cuales forman una secuencia anidada de subconjuntos concretos) es previsible que el aporte del conflicto al valor total de la ambigüedad sea mínimo. Parece entonces muy conveniente encontrar una manera de calcular la incertidumbre generada en el proceso de agrupación borrosa. Una forma sería la aplicación de medidas de incertidumbre asociadas con la vaguedad, y medidas de no-especificidad.

VI. UNA PROPUESTA PARA LA ESTIMACIÓN DE LA CALIDAD DE AGRUPAMIENTO BORROSO DE DATOS

El tema de la calidad del agrupamiento de datos es tratado, entre otras áreas disciplinares, en la estadística y la informática, donde se lo denomina *análisis de agrupamiento* o *clustering analysis*, como es conocido en inglés. Las referencias [26]-[29] pueden ser consultar para ahondar en el tema del análisis de agrupamiento.

Esta sección está dedicada a presentar parte del avance de una propuesta de evaluación de la calidad de agrupamiento de datos mediante CBMDs. Esta propuesta de evaluación se sustenta en la determinación de funciones matemáticas que ayuden a estimar la “calidad del agrupamiento” de datos a través de CBMDs, con base en dos aspectos: uno que da cuenta de la repartición o distribución de los datos del sistema en los CBMDs; el segundo alude a la congruencia o cohesión de cada CBMD con los datos que agrupa. Un buen proceso de agrupamiento debe producir CBMDs que cubran adecuadamente la región de los datos del sistema, al tiempo que en cada conjunto se ubiquen datos con características o comportamientos similares.

A. Calidad de agrupamiento por distribución de datos

Se asume que en cada CBMD A_i se calcula esta característica, simbolizada como Qd_i. La *distribución* de los datos en los CBMDs, está a su vez estrechamente relacionada con el *solapamiento relativo*, Sr_i, de los datos entre los CBMDs. Luego de establecer ciertos requerimientos matemáticos para cualquier expresión matemática que calcule Qd_i, con base en Sr_i, se ha propuesto la siguiente función:

$Qd_i = \frac{1}{Sr_i}$; donde Sr_i es el solapamiento relativo exhibido por el CBMD A_i.

Solapamiento relativo. Se define como la porción de solapamiento que le corresponde a cada CBMD A_i, y es calculada como:

$Sr_i = \frac{Sa_i}{\sum_{i=1}^{L} Sa_i}$; donde *L*, es el número de CBMDs; Sa_i es el *solapamiento absoluto* del CBMD A_i. El solapamiento, sea absoluto o relativo, informa sobre cómo fueron repartidos los elementos del espacio universo X en los diferentes CBMDs obtenidos en el proceso de agrupamiento de datos. Intuitivamente, un proceso de generación de CBMD se considera más afortunado, cuanto más diferenciados sean los CBMD descubiertos, es decir, menos "solapados sean".

Ahora, el solapamiento absoluto se determina como: $Sa_i = \sum_{j=1;\, j\neq i}^{L} S_{ij}$ en el que S_{ij} grado de solapamiento del CBMD A_i respecto del CBMD A_j; el cual puede estimarse a través de: $S_{ij} = \frac{|A_i \cap A_j|}{|A_j|}$; donde $A_i \cap A_j$: intersección estándar entre los CBMDs A_i y A_j; $|A_i|$: cardinalidad escalar del CBMD A_i definida como: $|A_i| = \sum_{x\in X} \mu_{A_i}(x)$ en la que $\mu_{A_i}(x)$ es el grado de pertenencia del elemento x en el CBMD A_i.

B. *Calidad de agrupamiento por cohesión de datos*

La calidad de agrupamiento por cohesión es una característica directamente asociada con la compacidad relativa de cada CBMD. Luego de establecer ciertos requerimientos axiomáticos para definir la relación entre la cohesión para cada CBMD A_i y su compacidad relativa; se ha propuesto la siguiente función:

$Qc_i = \exp(Cr_i) - 1$; donde Cr_i es la compacidad relativa del CBMD A_i.

Compacidad relativa. Se define como la porción de compacidad que le corresponde a cada CBMD A_i, y es calculada como:

$Cr_i = \frac{Ca_i}{\sum_{i=1}^{L} Ca_i}$; donde Ca_i es la compacidad absoluta del CBMD A_i. La compacidad, absoluta o relativa, se entiende como una medida de la cercanía de los elementos del conjunto a un dato que se reconoce como su *prototipo*. CBMDs densos o compactos indican que sus elementos exhiben naturaleza y comportamiento similares entre ellos y semejante al del prototipo del CBMD. La compacidad de un CBMD se expresará en términos de la proporción entre la cantidad de datos del conjunto (o sea su cardinalidad) y un valor que indique el tamaño de la región ocupada por esos datos.

$Ca_i = \frac{|A_i|}{\bar{d}_i^2}$; donde $|A_i|$ es la cardinalidad escalar del CBMD A_i, $\bar{d}_i^2$ es la diferencia cuadrática media de los elementos de X al prototipo del CBMD A_i, calculada como:

$\bar{d}_i^2 = \frac{\sum_{x\in X} \mu_{A_i}(x) d^2(x, v_i)}{|A_i|}$; donde $d(x, v_i)$ es una distancia métrica entre el elemento x y el prototipo del CBMD A_i, v_i. Se ha determinado que el prototipo de un CBMD sea su centroide. El centroide v_i se calcula como: $v_i = \frac{\sum_{x\in X} \mu_{A_i}(x) x}{|A_i|}$.

VII. EXPLORACIÓN DE ÍNDICES DE ESTIMACIÓN DE INCERTIDUMBRE Y CALIDAD DE AGRUPAMIENTO DE CBMDS EN UN SIB T-S

En este artículo se presentan varias funciones que se han planteado de acuerdo con fundamentos matemáticos de la Teoría de la medida y examinado desde el punto de vista práctico al ser utilizadas en el proceso de identificación de un SIB T-S para un sistema a pequeña escala de un intercambiador de calor caldereta-condensador. Este ensamblaje está localizado en el Laboratorio de Control de Procesos y Operaciones Unitarias, Facultad de Minas, Universidad Nacional de Colombia, Medellín. Cada función pretende integrar dos aspectos: la incertidumbre y la calidad de agrupamiento en los CBMDs de un SIB T-S. A continuación serán descritos los índices no sin antes establecer una estructura matemática que obedecen todos estos índices.

$I(m) = \sum_{i=1}^{L_m} I_i(m) = \sum_{i=1}^{L_m} u_i(m) q_i(m)$; donde:

$I(m)$: valor del índice en el CBMD A_i en el modelo m.

$I_i(m)$: valor del índice del modelo m.

L_m: es el número de CBMDs que aparecen en la partición borrosa m.

$u_i(m)$: es una función monótonamente creciente respecto a las componentes de incertidumbre del CBMD A_i en el modelo m, $u_i(m) = g(B_i(m), V_i(m))$.

$B_i(m)$: es la cantidad de ambigüedad del CBMD A_i en el modelo m, así, $B_i(m) = CF_i(m) + NE_i(m)$.

$CF_i(m)$: es la cantidad de conflicto del CBMD A_i en el modelo m.

$NE_i(m)$: es la cantidad de no-especificidad en el CBMD A_i en el modelo m.

$V_i(m)$: es la cantidad de vaguedad del CBMD A_i, en el modelo m.

$q_i(m)$: es una función monótonamente decreciente respecto de las componentes de la calidad del CBMD A_i en el modelo

m, $q_i(m) = h(Qc_i(m), Qd_i(m))$.

$Qc_i(m)$, valor de la calidad referida con la cohesión de los datos en el CBMD A_i en el modelo m.

$Qd_i(m)$: valor de la calidad referida con la distribución de los datos asignados en el CBMD A_i en el modelo m.

Estos dos últimos términos merecerán especial atención, en líneas posteriores; serán empleados sin revelar sus definiciones formales, momentáneamente.

Índice 1. $I_U(m) = \sum_{i=1}^{L_m} U_i(m)$,

Caso en que: $u_i(m) = U_i(m) = B_i(m) + V_i(m)$ y $q_i(m) = 1$.

Índices 2 $I_{U/Q+} = \sum_{i=1}^{Lm} \frac{U_i(m)}{Qc_i(m) + Qd_i(m)}$

Caso en que:

$u_i(m) = U_i(m) = B_i(m) + V_i(m)$;

$q_i(m) = \frac{1}{Q_i(m)}$ y $Q_i(m) = Qc_i(m) + Qd_i(m)$

Índice 3 $I_{U/Q*} = \sum_{i=1}^{Lm} \frac{B_i(m) + V_i(m)}{Qc_i(m) Qd_i(m)}$

Caso en que:

$u_i(m) = U_i(m) = B_i(m) + V_i(m)$;

$q_i(m) = \frac{1}{Q_i(m)}$ y $Q_i(m) = Qc_i(m) + Qd_i(m)$

Índice 4 $I_{U/Q_+ r} = \sum_{i=1}^{L_m} \frac{1/Q_i(m)}{\sum_{i=1}^{L_m} 1/Q_i(m)} U_i(m)$

Caso en que:

$u_i(m) = U_i(m) = B_i(m) + V_i(m)$;

$q_i(m) = \frac{1/Q_i(m)}{\sum_{i=1}^{L_m} 1/Q_i(m)}$ y $Q_i(m) = Qc_i(m) + Qd_i(m)$

Índice 5. $I_{U_+ r/Q_+ r} = \sum_{i=1}^{L_m} \frac{1/Q_i(m)}{\sum_{i=1}^{L_m} 1/Q_i(m)} \frac{U_i(m)}{\sum_{i=1}^{L_m} U_i(m)}$

Caso en que:

$u_i(m) = \frac{U_i(m)}{\sum_{i=1}^{L_m} U_i(m)}$; $q_i(m) = \frac{1/Q_i(m)}{\sum_{i=1}^{L_m} 1/Q_i(m)}$;

$U_i(m) = B_i(m) + V_i(m)$; $Q_i(m) = Qc_i(m) + Qd_i(m)$

VIII. ANÁLISIS DE RESULTADOS

En la actualidad no se poseen resultados definitivos que puedan ayudar a establecer un análisis confiable sobre la bondad de alguno(s) de los índices propuestos en la sección 7. Sólo ha podido establecerse que el índice 3 no es apropiado para su uso debido al comportamiento bastante restrictivo en caso que los valores de Qc y Qd sean menores a uno. Se dice restrictivo porque la interacción de ambos valores disminuye su efecto, en lugar de aumentarlo. Así entonces, de ahora en adelante cuando se haga referencia a los índices, se hablará de los índices 1, 2 4 y 5.

Sin embargo, algunos resultados preliminares han mostrado que los índices 1, 2 y 4 exhiben un comportamiento similar, mientras que el índice 5 presenta una conducta oscilatoria en una región donde las demás son monótonas. Otro resultado parcial interesante es aquel en el que se establece que cualquiera de los índices 1, 2 y 4 pueden servir como criterio evaluador de modelos intermedios o final; tarea que es regularmente efectuada por algún índice de desempeño, tal como el *error cuadrático medio*.

IX. CONCLUSIÓN

En este artículo se ha puesto de presente la posibilidad de emplear un índice de incertidumbre para evaluar y seleccionar un modelo, intermedio o final, en un proceso de identificación de SIB T-S. Este índice puede emplearse como criterio adjunto al tradicional índice de desempeño utilizado en estos procesos. Adicionalmente, se han expuesto diferentes fuentes de agregación de incertidumbre en un proceso de identificación de modelos y en particular cuando se emplea un método de identificación que genera CBMD en las reglas de los SIB T-S. Finalmente, ha sido importante presentar una propuesta de evaluación de una partición borrosa mediante los conceptos de calidad de agrupamiento borroso por distribución de datos y cohesión de CBMDs.

REFERENCIAS

[1] L. Ljung, System identification. Theory for the user. New Jersey: Prentice-Hall, 1987.

[2] G. J. Klir y T. A. Folger, Fuzzy sets, Uncertainty and Information. Englewood Cliffs, NJ: Prentice Hall, 1988.

[3] G. J. Klir y D. Elias, Architecture of Systems Problem Solving. International Federation for Systems Research. International Series on systems Science and Engineering, 2nd ed., vol. 21, Ed. Kluwer academic/Plenum publishers, 2003.

[4] G. J. Klir y M. J. Wierman, Uncertainty-Based Systems. Heidelberg: Physical Verlag, 1998.

[5] L. G. Parrat, Probability and Experimental Errors in Science, John Wiley, 1961.

[6] L.A. Zadeh, "Fuzzy Sets", Information and Control, vol. 8, pp. 338-353. June 1965.

[7] A. P. Dempster, "Upper and lower probabilities induced by multivalued mappings", en: Annals of Mathematical Statistics, vol. 38, pp. 325-339. 1967.

[8] G. Shafer, A Mathematical Theory of Evidence. Princeton University Press, 1976.

[9] L.A. Zadeh, "Fuzzy Sets as a Basis for a Theory of Possibility", Fuzzy Sets and Systems, vol.1, no.1, pp.3-28, 1978.

[10] D. Dubois and H. Prade, Possibility Theory: An approach to Computerized Processing of Uncertainty. 1st ed., New York: Plenum Press, 1988.

[11] P. Walley, Statistical Reasoning with Imprecise Probabilities. London: Chapman and Hall, 1991.

[12] G. J. Klir and R. M. Smith, "Recent Developments in Generalized Information theory", International Journal of Fuzzy Systems, vol. 1, no. 1, Sept. 1999.

[13] R.V.L. Hartley, "Transmission of Information", The Bell System Technical Journal, vol. 7, pp. 535-563, 1928.

[14] , M. Higashi y G. J. Klir, "Measures of Uncertainty and Information Based on Possibility Distributions", International Journal of General Systems, vol. 9, no.1, pp. 43-58, 1983.

[15] D. Dubois y H. Prade, "A note on Measures of Specificity for Fuzzy Sets", International Journal of General Systems, vol. 10, no 4, pp. 291-371, 1985.

[16] J. Abellán y S. Moral, "A Non-Specificity Measure for Convex Sets of probability Distributions", en: Proc. First Intl. Symposium on Imprecise probabilities and their Applications, Gent, Belgium, 1999.

[17] C. Shannon, "The Mathematical Theory of Communication", The Bell System Technical Journal, vol. 27, pp.379-423; 623-658, 1948.

[18] A. DeLuca y S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory", Information and Control, vol. 20, no 4, pp. 301-312, 1972.

[19] R. R. Yager, "On the Measure of Fuzziness and Negation. Part I: Membership in the Unit Interval", International Journal of General Systems, vol. 5., pp. 221-229, 1979.

[20] A. Kaufmann, Introduction to the Theory of Fuzzy Subsets. Volumen I Fundamental Theoretical Elements. New York: Academic Press, 1975.

[21] T. Takagi y M. Sugeno, "Fuzzy Identification of systems and its application to modeling an control", IEEE Transactions on Systems Man and Cybernetics, vol. 15, no. 1, pp. 116-132, 1985.

[22] M. Peña, "Contenido de Información en la identificación de sistemas", Instituto de Control Automático, Universidad Nacional de San Juan, Argentina, comunicación privada, 2005.

[23] M. Peña, F. di Sciascio y R. Carelli, "Identificación de la estructura de un modelo borroso del tipo Takagi–Sugeno", en: VIII Congreso Latinoamericano de Control Automático, Viña del Mar, Chile, vol. 2, pp. 403-408, 1998.

[24] M. Peña, Fuzzy model based control. (In Spanish). Ph.D. Thesis. INAUT, Facultad de Ingeniería, Universidad Nacional de San Juan, ISBN 950-605-278-6. EFU, San Juan, Argentina, 2001.

[25] M. Peña, y H. Álvarez, "Implicit Multidimensional Fuzzy Sets in Takagi-Sugeno Fuzzy Modeling", en: Congreso internacional en Inteligencia Computacional, Medellín, Colombia, 2001.

[26] S. Everitt, S. Landau, y M. Leese, Custer Analysis. A Hodder Arnold Publication, 4a ed., 2001.

[27] M. Halkidi, Y. Batistakis y M. Vazirgiannis, "Clustering algorithms and validity measures", en: Thirteenth International Conference on Scientific and Statistical Database Management, Fairfax, Virginia, July 18-20, 2001.

[28] L. Kaufman, y P. J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis (Wiley Series in Probability and Statistics), Wiley-Interscience; 2 Rev Ed. ed., 2005.

[29] B. Mirkin, Clustering for Data Mining (Computer Science and Data Analysis). Chapman & Hall/CRC, 2005.

Carlos Mario Sierra Se desempeña como docente de tiempo completo del Programa de Ing. de Sistemas de la Universidad de Antioquia. Hace parte del grupo de investigación de Simulación de Comportamientos de Sistemas (SICOSIS), adscrito a la facultad de Ingeniería de la misma Universidad. Titulado como Ingeniero civil y Master en Ingeniería de Sistemas de la Universidad Nacional de Colombia Sede Medellín. Actualmente candidato a Doctor en el programa de Doctorado en Ingeniería, énfasis Sistemas e Informática, de la Universidad Nacional Sede Medellín. Las áreas de trabajo son fundamentalmente, el tratamiento matemático de la incertidumbre y los sistemas basados en conocimiento.

Universidad Nacional de Colombia Sede Medellín
Facultad de Minas

120 años
TRABAJO Y RECTITUD

Escuela de Ingeniería de Sistemas

Pregrado

- Ingeniería de Sistemas e Informática.

Posgrado

- Doctorado en Ingeniería-Sistemas.
- Maestría en Ingeniería de Sistemas.
- Especialización en Sistemas con énfasis en:
  - Ingeniería de Software.
  - Investigación de Operaciones.
  - Inteligencia Artificial.
- Especialización en Mercados de Energía.

Áreas de Investigación

- Ingeniería de Software.
- Investigación de Operaciones.
- Inteligencia Artificial.

Escuela de Ingeniería de Sistemas
Dirección Postal:
Carrera 80 No. 65 - 223 Bloque M8A
Facultad de Minas. Medellín - Colombia
Tel: (574) 4255350 Fax: (574) 4255365
Email: esistema@unalmed.edu.co
http://pisis.unalmed.edu.co/