

# Toma de decisiones mediante técnicas de razonamiento incierto

José Fernando Reyes Saldaña

Programa de Postgrado en Ingeniería de Sistemas FIME-UANL

fernando@yalma.fime.uanl.mx

Rodolfo García Flores

Sistemas Lógicos SisLogic, S.A.

garciaflores.r@gmail.com

## RESUMEN

*En el mundo real no existe algo que sea ciento por ciento seguro. Todos los días nos topamos con situaciones que implican el tomar una decisión con información imprecisa. El método que se ha utilizado tradicionalmente para manejar estos problemas ha sido el modelo de Bayes. Buscando algunas alternativas que sobrepasaran las limitaciones mostradas por este método, se han creado otros modelos que tratan ciertos tipos de incertidumbre, cada uno de una manera diferente. Entre estos métodos está la teoría de la lógica difusa, creada por Lofti Zadeh y que trata la vaguedad en la información. Otro método utilizado es la teoría de la evidencia de Dempster y Shafer. Esta teoría trata la ambigüedad en la información, y provee una gran variedad de medidas que dan información más precisa acerca del tipo de incertidumbre en los datos.*

*En este artículo se mostrarán las ventajas y desventajas de cada uno de los métodos mencionados, se explica por qué se considera la teoría de la evidencia de Dempster-Shafer una de las que tiene mayor potencial, y se presenta la biblioteca numérica LIDSET para el manejo de incertidumbre.*

## PALABRAS CLAVE

Incertidumbre, modelo bayesiano, lógica difusa, teoría de la evidencia, Dempster-Shafer.

## ABSTRACT

*In the real world nothing is one hundred per cent sure. Every day we find situations that imply taking decisions with uncertain information. The method used traditionally to solve these problems has been the Bayesian model. Looking for alternatives to take over the limitations of this method, several models have been created that manage different kinds of uncertainty, each following a different approach. One of those methods is fuzzy logic theory, created by Lofti Zadeh, for modelling vagueness in information. A different method is evidence theory, created by Glenn Shaffer. This theory takes ambiguity on information, and has several information measures akin to entropy that give a better insight about uncertainty.*

*In this paper we show the advantages and disadvantages of each of these methods, explain why we consider that Dempster-Shafer's evidence theory*



*offers the greatest potential advantage, and present the numerical library LIDSET for uncertainty management.*

## KEYWORDS

Uncertainty, Bayesian model, fuzzy logic, evidence theory, Dempster-Shafer.

## INTRODUCCIÓN

El crear conclusiones útiles a partir de información incompleta o imprecisa no es una tarea imposible, pues los seres humanos lo hacemos casi en cada aspecto de nuestra vida diaria. Los médicos realizan diagnósticos correctos y recomiendan tratamiento a partir de síntomas ambiguos, los mecánicos analizan los problemas de los automóviles a partir de observaciones, y todos los seres humanos comprendemos el lenguaje hablado, corporal o escrito a pesar de tener frases incompletas y reconocemos a otras personas por sus voces o sus gestos.<sup>1</sup>

Así, para realizar el proceso de toma de decisiones se utiliza la lógica. Con lógica algunas piezas de conocimiento son utilizadas en el razonamiento, y pueden ser parte de las explicaciones o conclusiones. Desde los años sesentas del siglo pasado se ha tratado de automatizar la toma de decisiones en diversos campos del conocimiento a través de sistemas expertos basados en reglas lógicas de la forma “Si... entonces...” con resultados muy positivos. Sin embargo, el empleo de la lógica tradicional en el desarrollo de sistemas de razonamiento automático tiene sus limitantes, especialmente en los casos en los que hay información faltante o incertidumbre, y en estos casos los procedimientos de inferencia tradicional pueden no ser útiles. Por ejemplo, puede suceder que las observaciones tengan un amplio margen de error, que las relaciones de causa a efecto no sean claras, que se requiera conocimiento no explícito u observable para tomar una decisión, o simplemente que los términos utilizados en la descripción del problema sean vagos o ambiguos. Para compensar estas deficiencias, se han propuesto diversas metodologías que han probado su valor al emplearse en sistemas de apoyo a la toma de decisiones.

El propósito de este artículo es presentar las principales aproximaciones matemáticas propuestas

para atacar el problema de razonamiento con incertidumbre, poniendo énfasis en la teoría de la evidencia de Dempster-Shafer. A continuación se presenta el enfoque probabilístico clásico (o bayesiano), que es actualmente el más desarrollado y utilizado. En la sección siguiente se explican los conceptos básicos de la lógica difusa con ejemplos ilustrativos de su uso. Después se presenta la biblioteca numérica LIDSET, desarrollada por los autores para el estudio de la incertidumbre. Ella se utilizará para introducir los fundamentos de la teoría de la evidencia de Dempster-Shafer junto con sus ventajas, aplicaciones y problemas abiertos a la investigación. Para concluir, los comentarios finales resumen las ventajas y problemas que presentan en su uso los métodos descritos.

## LA PROBABILIDAD CLÁSICA

El enfoque preferido para el tratamiento de la incertidumbre ha sido hasta el día de hoy el modelo de Bayes, aunque éste tiene varias desventajas, que se explicarán más adelante. El modelo bayesiano se define por los siguientes tres atributos:<sup>2</sup>

- Depende de un modelo probabilístico completo, definiendo probabilidades para todos y cada uno de los eventos que formarán nuestro campo de estudio.
- Puede aceptar juicios subjetivos cuando no se tiene información completa de datos empíricos para definir probabilidades.
- Utiliza el teorema de Bayes como el mecanismo primordial para la actualización de las credibilidades en cuanto se obtiene información nueva.

El teorema de Bayes, utilizado para la actualización de las credibilidades en probabilidad clásica, se define por la ecuación (1).

$$P(F_j|E) = \frac{P(EF_j)}{P(E)} = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^n P(E|F_i)P(F_i)} \quad (1)$$

Para detalles sobre su deducción, el lector puede consultar las referencias.<sup>3 o 4</sup>

La ecuación (1) muestra la forma de calcular la probabilidad de que el evento  $F_j$  ocurra dado que el evento  $E$  acaba de ocurrir, suponiendo que los elementos del conjunto  $F_i$  son eventos mutuamente

excluyentes (es decir, no pueden ocurrir dos de estos eventos al mismo tiempo). En otras palabras, podemos considerar a  $E$  como evidencia de  $F_j$ , y calcular la probabilidad de que  $F_j$  ocurra dada la evidencia,  $E$ . Para ilustrar la utilidad del teorema de Bayes, supongamos un caso simple en el que sólo existen los eventos  $F$  y  $F'$ , i.e.  $j = 2$ . Sea el evento  $F$  el padecer meningitis, y sea  $E$  el padecer migraña. La ecuación (1) permitiría calcular la probabilidad de padecer meningitis dado que se tiene migraña (un conteo muy difícil entre la población) a partir del conocimiento del número de casos que tienen migraña y que también han presentado la meningitis (un conteo muy fácil de hacer en un hospital), y la probabilidad de padecer meningitis en general.

La utilización de la teoría de la probabilidad supone la existencia de una magnitud (la *probabilidad*) que se asigna a la ocurrencia de un evento. Se supone que esta magnitud existe, independientemente de que el observador-tomador de decisiones tenga la posibilidad de conocerla o no.

Idealmente, el observador podría asignar siempre la magnitud correspondiente a un evento dado si pudiera conocer el número de casos en que ese evento ocurre de entre el número de experimentos totales. Mientras mayor fuera el número total de experimentos, mejor sería la aproximación de esa magnitud al valor real de probabilidad. Sin embargo, no siempre es posible contar casos (*enfoque frecuentista*) y muchas veces el tomador de decisiones debe manejar la maquinaria teórica comenzando simplemente con un estimado de los valores de probabilidad (el llamado *enfoque subjetivo*).

Se ha argumentado<sup>2</sup> que si el tomador de decisiones se ve obligado a utilizar estimados, entonces también es válido emplear otras técnicas diferentes a la probabilidad, pues todo el marco teórico de la probabilidad clásica se justifica en el conocimiento exacto de los valores de probabilidad. Estas técnicas se explicarán en las siguientes secciones.

Recientemente,<sup>5</sup> se han propuesto nuevas ideas para enriquecer la teoría de la probabilidad actualizando el conocimiento que el tomador de decisiones adquiere a través de su experiencia. El resultado son las redes bayesianas, las cuales se explican a continuación.

## REDES BAYESIANAS

El uso de las redes bayesianas se ilustrará mediante un ejemplo clásico de la literatura.<sup>6</sup> El inspector de policía Smith está esperando impacientemente la llegada del Sr. Holmes y del Dr. Watson, quienes llegarán manejando sus respectivos autos. Holmes y Watson están retrasados y el inspector Smith tiene una cita importante. Hace frío y, mirando afuera por la ventana, Smith se pregunta si las carreteras están congeladas. Ambos son malos manejando, así que si la carretera está congelada lo más probable es que ambos hayan tenido algún accidente.

La secretaria entra y le dice a Smith que el Dr. Watson ha tenido un accidente en su automóvil. “Si Watson ha tenido un accidente, debe haber hielo en la carretera, y por consiguiente, Holmes debe haber chocado también. Ya me puedo ir a almorzar”.

“¿Hielo?”, responde la secretaria, “no hace tanto frío afuera, además, se ha puesto sal en las carreteras, no puede haber hielo”. El Inspector Smith reflexiona, “Watson tuvo mala suerte, vamos a darle a Holmes otros diez minutos”.

Para formalizar la historia se asignarán variables a los eventos ocurridos, con dos estados cada una, *sí* o *no*. Se supondrá también que cada evento tiene asociada cierta certidumbre, la cual se representa con un número real. Así, se tienen tres variables: hielo en la carretera ( $I$ ), Holmes choca ( $H$ ) y Watson choca ( $W$ ). La relación entre las variables se muestra en la figura 1. Así, al suceder  $I$  se incrementa la certeza de que  $H$  y  $W$  puedan suceder. En otras palabras, los choques están en función creciente de la certeza que se tenga sobre su causa.

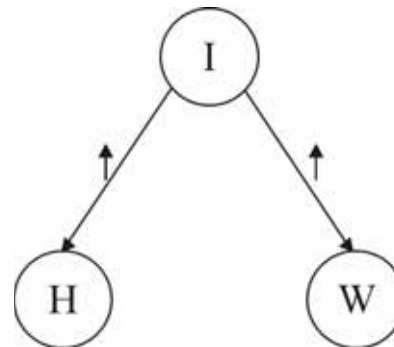


Fig. 1. Un modelo de red para el ejemplo de hielo en la carretera. Las flechas largas representan el impacto causal, las flechas pequeñas representan la dirección del impacto en la certidumbre.

Cuando al inspector Smith se le dice que Watson ha tenido un accidente en su automóvil, él hace un razonamiento en la dirección opuesta a las flechas causales. Dado que, al pasar el tiempo, la evidencia de impacto apuntando a  $W$  se incrementa, Smith asigna una mayor certidumbre a  $I$ . La certidumbre en  $I$  crea a su vez una nueva expectativa: mayor certidumbre en  $H$ . Luego, cuando la secretaria le dice a Watson que las carreteras no pueden estar congeladas, el hecho de que Watson haya tenido un accidente no cambia su certidumbre respecto a las condiciones de la carretera, y por consiguiente,  $W$  no tiene influencia sobre  $H$ .

Así, para calcular las probabilidades se necesita tener  $P(H|I)$ ,  $P(W|I)$  y  $P(I)$ . Esto simplifica las relaciones entre las variables en la ecuación (1), pues el usar la regla de Bayes con todas sus dependencias implicaría también el conocimiento de  $P(H|W)$ ; al incrementarse el número de variables en el sistema observado, la simplificación sería más significativa. La red bayesiana permite la actualización de los valores conforme se incrementa la experiencia de Smith; Jesen muestra los detalles en la implementación.<sup>6</sup> El modelo de la figura 1 muestra que solamente el conocimiento en  $I$  tiene efecto directo sobre  $H$  y  $W$ . Así, se debe asignar una certidumbre a  $I$  basada en el conocimiento disponible.

Las redes bayesianas resuelven el problema de las observaciones sucesivas hechas por el tomador de decisiones. Sin embargo, el enfoque bayesiano presenta otros problemas, como el ya mencionado abuso de la teoría cuando los valores de probabilidad son estimados subjetivos, y la falta de distinción entre los conceptos de ignorancia (falta de información) e incertidumbre (ambigüedad), pues ambos conceptos afectan la magnitud de la probabilidad. Los enfoques que se presentan a continuación atacan estos problemas.

## LÓGICA DIFUSA

Existen dos proposiciones que son esenciales para el uso de la lógica formal tradicional. La primera es con respecto a la membresía: para cualquier elemento y un conjunto perteneciente a cualquier universo, el elemento es miembro del conjunto o en caso contrario es miembro del conjunto complemento.

La segunda proposición afirma que un elemento no puede pertenecer al mismo tiempo a un conjunto dado y a su complemento. Ambas proposiciones son violadas en la teoría de conjuntos difusos de Lotfi Zadeh.<sup>7</sup> A las leyes de conjuntos y razonamiento de la lógica tradicional se les llama *estrictas* desde el punto de vista de los conjuntos difusos.<sup>1</sup>

La teoría de conjuntos difusos describe las propiedades de los conjuntos difusos, una clase de objetos con grados de membresía continuos en el intervalo  $[0,1]$ . Esta idea contrasta con la teoría de probabilidad tradicional, en la cual los objetos tienen un valor de membresía tomados del conjunto  $\{0,1\}$ . Cada objeto  $x$  en el conjunto difuso  $X$  tiene asignado un grado de membresía a un conjunto  $Y \in X$  dado por una función usualmente denotada por  $\mu(x)$ , cuyos valores se encuentran entre 0 y 1. Frecuentemente se confunde la idea de una función de membresía  $\mu(x)$  con una función de densidad de probabilidad  $f(x)$ , sin embargo la integral de  $f(x)$  debe sumar 1, mientras  $\mu(x)$  no tiene esa restricción.<sup>8</sup>

La lógica difusa maneja el concepto de la vaguedad en la evidencia. El lenguaje natural está lleno de conceptos vagos e imprecisos, como por ejemplo decir “Juan es alto” o “El día está caliente”. Estas oraciones son difíciles de traducir en un lenguaje más preciso sin perder su valor semántico. Por ejemplo, decir “Juan mide 170 cm” no indica que Juan sea alto. Por ello, para añadir precisión es necesario buscar otra forma de formularlas.

Existen muchas clases de objetos en el mundo real que no tienen un criterio definido para describirlas como pertenecientes a un grupo u otro, por ejemplo, la clase de automóviles caros, la clase de ingenieros mal pagados, etc. Sin embargo, este tipo de categorías sin una definición precisa juegan un papel importante en el pensamiento humano, particularmente en el dominio de reconocimiento de patrones, comunicación de información, toma de decisiones y abstracción. La teoría de conjuntos difusos ha sido muy utilizada para manejar este problema. Por ejemplo, los japoneses han sido pioneros en esta rama aplicando teoría de control difuso a cualquier cosa, desde procesos de control industrial, operación de trenes automáticos, reconocimiento de habla e imágenes, horarios de autobuses, sistemas de toma de decisiones, diagnóstico médico, programación

multi-objetivos, elevadores, enfocado en cámaras de video y manufactura de automóviles.

Una de las mejores explicaciones en cuanto a cómo la teoría de conjuntos difusos es auxiliar en la toma de decisiones es dado por Yager<sup>9</sup>: “Debemos de tener en mente que el uso de conjuntos difusos no elimina la naturaleza subjetiva o difusa de los conceptos con los que trabajamos... pero nos da una manera de manejar los conceptos subjetivos en una forma racional, en una forma similar a cómo el método utilizado por los tomadores de decisiones bayesianos les permite manejar probabilidades subjetivas”.

Ejemplos de conjuntos difusos se presentan en la figura 2 y figura 3.<sup>1</sup> En la figura 2,  $S$  es el conjunto de enteros positivos y  $F$  el subconjunto difuso de  $S$  llamado enteros pequeños. Ahora, varios valores enteros pueden tener la distribución de “posibilidad” definiendo su función de membresía en el conjunto de enteros pequeños:  $\mu(1) = 1.0$ ,  $\mu(2) = 1.0$ ,  $\mu(3)$

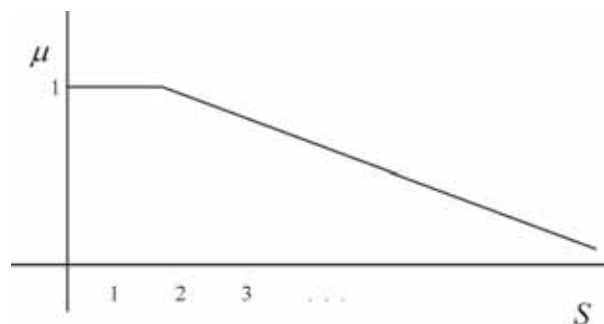


Fig. 2. La representación del conjunto difuso para enteros pequeños.

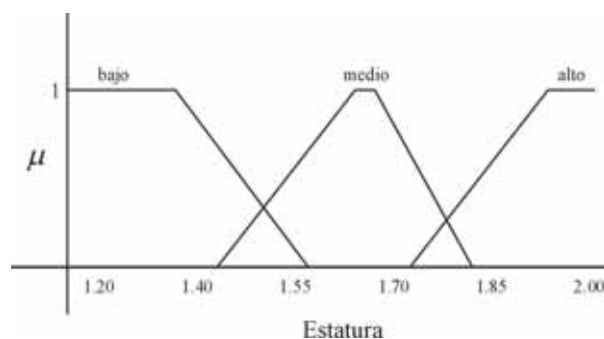


Fig. 3. Una representación difusa para el conjunto de estaturas para una población masculina, describiéndolos como bajo, medio, alto.

$= 0.9$ ,  $\mu(4) = 0.8$ , ...,  $\mu(50) = 0.001$ , etc. Para la proposición “el entero positivo  $X$  es un entero pequeño”, crea una distribución de posibilidad en todos los enteros positivos ( $S$ ).

En la figura 3 se muestra una función de membresía para el concepto de estatura de una población masculina, definiéndolos como bajo, medio y alto. Hay que notar que cualquier persona puede pertenecer a más de un conjunto, por ejemplo alguien que mide 1.77 m pertenece tanto al conjunto de medio como al de alto.

La teoría de los conjuntos difusos no se preocupa por cómo estas distribuciones de posibilidad son creadas, sino en las reglas de cómo se calcularán las posibilidades combinadas cuando se combinan varias expresiones que contienen variables difusas.

Así, la teoría de conjuntos difusos permite manejar la vaguedad en la evidencia, la cual se define como la dificultad al hacer una distinción precisa entre varios elementos del mundo real, es decir, el no poder definir de forma precisa el grupo al que pertenece cierto elemento, en contraste con la teoría de la evidencia, que maneja la ambigüedad, definida como la situación en la que cualquiera de dos elecciones es totalmente válida para la clasificación de un elemento.<sup>10</sup> Enseguida se describirá el marco teórico de la teoría de la evidencia y el concepto de ambigüedad.

## TEORÍA DE LA EVIDENCIA

La teoría de Dempster-Shafer<sup>11</sup> ha sido propuesta como un método para describir incertidumbre en la evidencia. Esta teoría se centra en la credibilidad asignada a que un evento pueda ocurrir (o haya ocurrido), desde el punto de vista y de acuerdo a la experiencia del tomador de decisiones, en contraste con la probabilidad clásica, que supone la existencia de valores de probabilidad asociados a eventos determinados independientemente de que el observador pueda conocer el valor real de la probabilidad. DS permite que la evidencia adquirida mediante observaciones o experimentos apoye al mismo tiempo varias conclusiones mutuamente excluyentes o ninguna conclusión en particular.

Para facilitar la investigación y el desarrollo de aplicaciones utilizando los conceptos de DS, los

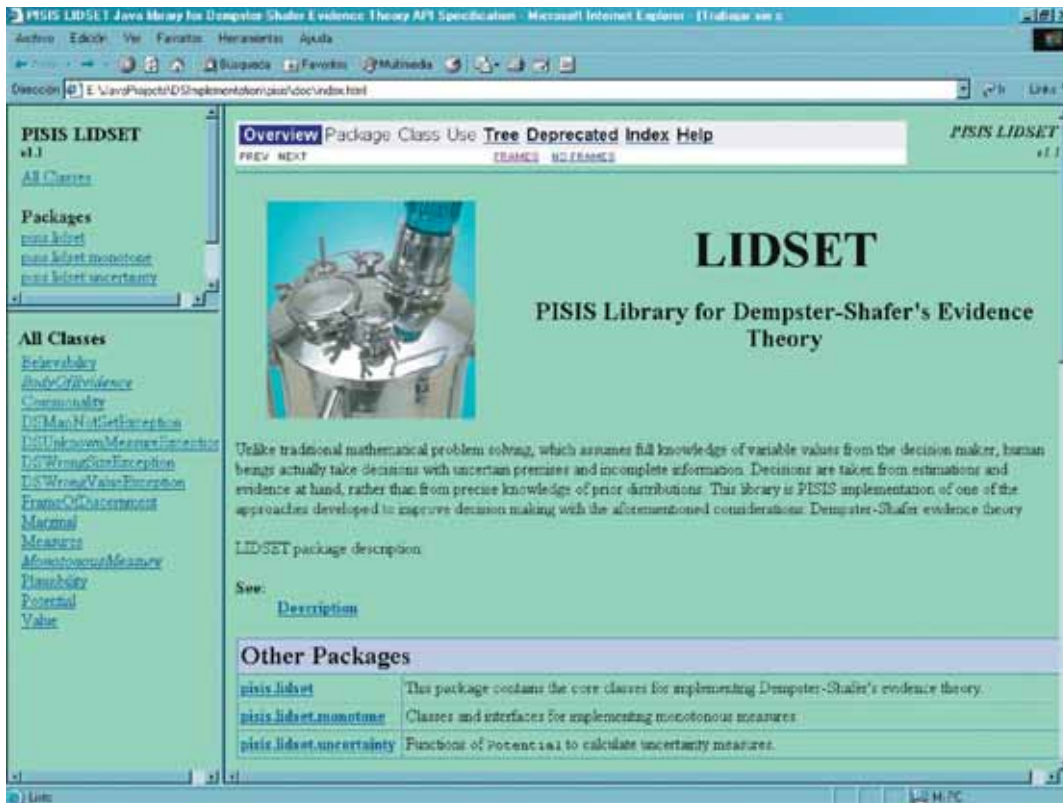


Fig. 4. Interface de protocolo de aplicación de la biblioteca numérica LISSET

autores han desarrollado la biblioteca numérica LISSET, cuya interface de protocolo de aplicación (API) se muestra en la figura 4. LISSET es una biblioteca de funciones programada en lenguaje Java totalmente orientada a objetos, portable e independiente de plataforma. Los resultados de los ejemplos que se presentan a continuación fueron obtenidos utilizando LISSET.

### Principios de la teoría de Dempster-Shafer

DS nos permite representar la imprecisión e incertidumbre por medio de la definición de dos funciones: plausibilidad ( $[\varphi]_p$ ) y credibilidad ( $[\varphi]_b$ ), derivadas de una asignación básica ( $[\varphi]_m$ ).

Se utilizará la notación propuesta por Haenni et al.<sup>12</sup> debido a su generalidad y conveniencia. Tendremos una pieza de evidencia (el potencial de credibilidad  $\varphi$ ), la cual puede tener varias representaciones, como son la asignación básica  $[\varphi]_m$ , la plausibilidad  $[\varphi]_p$ , la credibilidad  $[\varphi]_b$ , etc., que se definirán a continuación.

El *potencial de credibilidad*  $\varphi$  se define en un conjunto finito de variables  $D = \{x_1, x_2, \dots, x_n\}$ , el cual

es llamado *dominio de  $\varphi$* . Cada variable  $x_i$  tiene un conjunto correspondiente  $\Theta_{x_i}$  de valores posibles. El producto cartesiano  $\Theta_D = \Theta_{x_1} \times \dots \times \Theta_{x_n}$  es el conjunto de todos los valores posibles de  $D$ , y es llamado el *marco de discernimiento* de  $\varphi$ . Si no ha sido especificado  $D$ , al dominio de  $\varphi$  se le denotará como  $d(\varphi)$ .

La función de asignación básica  $[\varphi]_m : 2^{\Theta_D} \rightarrow [0,1]$  en  $D$  asigna a cada conjunto  $A \subset \Theta_D$  un valor en el intervalo  $[0,1]$ , el cual se denotará como  $[\varphi(A)]_m$ . Usualmente la siguiente condición debe de cumplirse:

$$\sum_{A \in \Theta_D} [\varphi(A)]_m = 1 \quad (2)$$

Los potenciales de credibilidad para los cuales esta ecuación se cumple se llaman *completos*. Denotaremos por  $\Phi_D^C$  al conjunto de todos los potenciales completos en  $D$ .

Algunas veces se impone otra condición,  $[\varphi(\emptyset)]_m = 0$ . Un potencial de credibilidad para el cual esta condición se cumple se llama *normalizado*,

de otro modo  $\varphi$  se llama *no normalizado* y  $c_\varphi = [\varphi(\emptyset)]_m$  es la *masa en conflicto*.

La asignación básica  $[\varphi(A)]_m$  indica el grado de evidencia que respalda la afirmación que un elemento específico de  $\Theta_D$  pertenezca a  $A$ , pero no a un subconjunto en particular.

Los conjuntos  $A \subseteq \Theta_D$  para los cuales  $[\varphi(A)]_m \neq 0$  son llamados *elementos focales* o *conjuntos focales*.  $FS(\varphi)$  denota el conjunto de todos los conjuntos focales de  $\varphi$  y  $FS'(\varphi) = FS(\varphi) \setminus \{\emptyset\}$  el conjunto correspondiente de todos los conjuntos focales no vacíos.

Más aún:

$$|\varphi| := |FS'(\varphi)| \tag{3}$$

define el tamaño de  $\varphi$ . Un potencial  $\varphi$  es representado usualmente por el conjunto  $\{(F_1, m_1), \dots, (F_k, m_k)\}$  de todas las parejas  $(F_i, m_i)$  con  $F_i \in FS(\varphi)$  y  $m_i = [\varphi(F_i)]_m$ .

Las funciones de credibilidad  $[\varphi]_b : 2^{\Theta_D} \rightarrow [0,1]$  pueden ser definidas de acuerdo a su función de masa correspondiente como:

$$[\varphi(A)]_b := \sum_{\emptyset \neq B \subseteq A} [\varphi(B)]_m = \sum_{\substack{B \subseteq A \\ B \in FS'(\varphi)}} [\varphi(B)]_m \tag{4}$$

lo cual implica que  $[\varphi(\emptyset)]_b = 0$ .

La credibilidad  $[\varphi(A)]_b$  se interpreta como el grado de credibilidad que se tiene en que un elemento de  $\Theta_D$  se encuentre en el conjunto  $A$ .

Además, la función de plausibilidad  $[\varphi]_p : 2^{\Theta_D} \rightarrow [0,1]$  en  $D$  está definida por

$$[\varphi(A)]_p := \sum_{B \cap A \neq \emptyset} [\varphi(B)]_m = \sum_{\substack{B \cap A \neq \emptyset \\ B \in FS'(\varphi)}} [\varphi(B)]_m \tag{5}$$

esto implica que  $[\varphi(\emptyset)]_p = 0$  y  $[\varphi(A)]_b \leq [\varphi(A)]_p$  para todos los subconjuntos  $A \subseteq \Theta_D$ .

Además de contar con la siguiente propiedad

$$[\varphi(A)]_p = 1 - [\varphi(\bar{A})]_b \tag{6}$$

**Ejemplo 1.** Se tiene  $D = \{x\}$  como el dominio para un potencial de credibilidad  $\varphi \in \Phi_D$  con  $\Theta_D = \Theta_x = \{x_1, x_2, x_3\}$  y  $FS(\varphi) = \{\emptyset, \{x_1\}, \{x_3\}, \{x_1, x_2\}, \{x_2, x_3\}\}$ . Por ejemplo,  $x_1, x_2$  y  $x_3$  representan respectivamente el resultado de un partido de fútbol, siendo  $x_1$  que el equipo local gane,  $x_2$  que el equipo visitante gane y  $x_3$  que el partido quede empatado. Dadas las asignaciones básicas  $[\varphi(A)]_m$  de la tabla I (provenientes tal vez del juicio de un experto), se puede calcular su credibilidad y plausibilidad por



Tabla I. Tabla de asignación básica, credibilidad y plausibilidad para un potencial no normalizado.

Conjunto focal → Representación ↓	$\emptyset$	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$	$\{x_1, x_2, x_3\}$
$[\varphi(A)]_m$	0.2	0.2	0	0.1	0.2	0	0.3	0
$[\varphi(A)]_b$	0	0.2	0	0.1	0.4	0.3	0.4	0.8
$[\varphi(A)]_p$	0	0.4	0.5	0.4	0.7	0.8	0.6	0.8

medio de las fórmulas (4) y (5) y que aparecen también en la tabla. Por ejemplo, para el cálculo de  $[\varphi(\{x_1, x_2\})]_b$ ,

$$[\varphi(\{x_1, x_2\})]_b = [\varphi(\{x_1\})]_m + [\varphi(\{x_2\})]_m + [\varphi(\{x_1, x_2\})]_m = 0.2 + 0 + 0.2$$

$$[\varphi(\{x_1, x_2\})]_b = 0.4$$

que representa el grado de credibilidad (dada la evidencia de la tabla I) que apoya las afirmaciones  $x_1$  o  $x_2$ .

y  $[\varphi(\{x_1, x_2\})]_p$ ,

$$[\varphi(\{x_1, x_2\})]_p = [\varphi(\{x_1\})]_m + [\varphi(\{x_2\})]_m + [\varphi(\{x_1, x_2\})]_m + [\varphi(\{x_1, x_3\})]_m + [\varphi(\{x_2, x_3\})]_m + [\varphi(\{x_1, x_2, x_3\})]_m = 0.2 + 0 + 0.2 + 0 + 0.3 + 0$$

$$[\varphi(\{x_1, x_2\})]_p = 0.7$$

Que puede interpretarse como el grado hasta el cual no podemos dudar de las afirmaciones  $x_1$  y  $x_2$ . Nótese que el experto asigna un valor al conjunto vacío, lo que significa que asigna posibilidad de ser a una afirmación no especificada en el dominio.

Ahora bien, la tabla I muestra un potencial no normalizado. Para obtener el potencial normalizado se debe distribuir la proporción de masa en conflicto  $c_\varphi$  entre todos los elementos focales  $FS^*(\varphi)$ , como sigue:

Asignación básica normalizada

$$[\varphi(A)]_M := \begin{cases} 0, & \text{si } A = \emptyset \\ \frac{[\varphi(A)]_m}{1 - c_\varphi}, & \text{otro} \end{cases} \quad (7)$$

Y para la credibilidad y plausibilidad, se tiene

$$[\varphi(A)]_B := \sum_{B \subseteq A} [\varphi(B)]_M = \sum_{\substack{B \subseteq A \\ B \in FS(\varphi)}} [\varphi(B)]_M = \frac{[\varphi(A)]_b}{1 - c_\varphi} \quad (8)$$

$$[\varphi(A)]_P := \sum_{B \cap A \neq \emptyset} [\varphi(B)]_M = \sum_{\substack{B \cap A \neq \emptyset \\ B \in FS(\varphi)}} [\varphi(B)]_M = \frac{[\varphi(A)]_p}{1 - c_\varphi} \quad (9)$$

denotando con subíndices en letras mayúsculas los valores normalizados y subíndices en letras minúsculas los valores originales.

De esta manera, la tabla II se obtiene normalizando los datos de la tabla I.

Aquí se debe notar que  $[\varphi(\emptyset)]_B = [\varphi(\emptyset)]_P = 0$  y que  $[\varphi(\Theta_D)]_B = [\varphi(\Theta_D)]_P = 1$ . Además, ya se sabe que  $[\varphi(A)]_B \leq [\varphi(A)]_P$  para todos los subconjuntos las  $A \subseteq \Theta_D$ .

De otra manera, la normalización se puede definir como una transformación  $v: \Phi_D \rightarrow \Phi_D$  de un potencial no normalizado  $\varphi \in \Phi_D$  a un potencial normalizado  $v(\varphi) \in \Phi_D$  en la siguiente ecuación:

$$[v(\varphi)]_m := [\varphi]_M \quad (10)$$

Claro está que esto implica que  $[v(\varphi)]_b = [\varphi]_B$  y  $[v(\varphi)]_p = [\varphi]_P$ . Más aún,  $FS(v(\varphi)) = FS^*(v(\varphi)) = FS^*(\varphi)$ .

Las operaciones básicas para realizarse sobre potenciales de credibilidad son la *combinación* y *marginalización*. Combinación corresponde a la agregación de evidencia. Ésta toma dos potenciales  $\varphi_1 \in \Phi_{D_1}^C$  y  $\varphi_2 \in \Phi_{D_2}^C$  y produce un nuevo potencial  $\varphi_1 \otimes \varphi_2$  en el dominio  $D_1 \cup D_2$ . En esta fórmula se supone que  $\varphi_1$  y  $\varphi_2$  son potenciales definidos en el

Tabla II. Tabla de asignación básica, credibilidad y plausibilidad normalizados.

Conjunto focal → Representación ↓	$\emptyset$	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$	$\{x_1, x_2, x_3\}$
$[\varphi(A)]_M$	0	0.25	0	0.125	0.25	0	0.375	0
$[\varphi(A)]_B$	0	0.25	0	0.125	0.5	0.375	0.5	1
$[\varphi(A)]_P$	0	0.5	0.625	0.5	0.875	1	0.75	1



mismo dominio  $D$ . De otro modo,  $\varphi_1$  y  $\varphi_2$  pueden siempre ser extendidos al dominio  $D_1 \cup D_2$ . La combinación de  $\varphi_1$  y  $\varphi_2$  en el potencial  $\varphi_1 \otimes \varphi_2$  en  $D$  que se determina por:

$$\begin{aligned} [\varphi_1 \otimes \varphi_2(A)]_m &:= \sum_{B \cap C = A} [\varphi_1(B)]_m \cdot [\varphi_2(C)]_m \\ &= \sum_{\substack{B \cap C = A \\ B \in FS(\varphi_1), C \in FS(\varphi_2)}} [\varphi_1(B)]_m \cdot [\varphi_2(C)]_m \end{aligned} \quad (11)$$

Esta forma de combinar evidencia de dos potenciales diferentes se conoce como la *regla de combinación de Dempster*. Descansa en la suposición de que  $\varphi_1$  y  $\varphi_2$  representan dos piezas de evidencia independientes. Desde el punto de vista del cómputo de este resultado, combinar los potenciales de  $\varphi_1$  y  $\varphi_2$  significa el intersectar cada conjunto focal  $B \in FS(\varphi_1)$  con cada conjunto focal  $C \in FS(\varphi_2)$ , después multiplicar los valores correspondientes de  $[\varphi_1(B)]_m$  y  $[\varphi_2(C)]_m$ , y finalmente sumar los valores resultantes para las mismas intersecciones.

Ejemplo 2. La regla de combinación se ejemplificará con un diagnóstico médico. Se supone  $Q$  como el dominio del problema, con cuatro elementos focales: gripa ( $C$ ), resfrío ( $F$ ), migraña ( $H$ ) y meningitis ( $M$ ). La tarea será poner asignación básica a los elementos del dominio  $Q$ . Se puede asignar un valor a conjuntos de estas hipótesis, por ejemplo, el tener fiebre podría apoyar  $\{C, F, M\}$ . Dado que los elementos de  $Q$  son hipótesis mutuamente excluyentes, la evidencia a favor de una puede afectar la credibilidad hacia otra.



Se supone que la primera pieza de evidencia es que el paciente tiene fiebre, la cual tiene un apoyo de  $[\varphi_1(\{C, F, M\})]_m = 0.6$ . Si esta fuera la única hipótesis, entonces  $[\varphi_1(Q)]_m = 0.4$ , que indica que el resto del apoyo se reparte entre todos los elementos de  $Q$ .

Ahora se obtiene nueva evidencia para el diagnóstico, pues el paciente tiene mucha náusea, lo que nos sugiere  $[\varphi_2(\{C, F, H\})]_m = 0.7$  y  $[\varphi_2(Q)]_m = 0.3$ .

Ahora bien, con la regla de Dempster se intersectan los conjuntos focales de  $\varphi_1$  con  $\varphi_2$ , como se muestra en la tabla III, obteniendo un nuevo potencial  $\varphi_3$  con las asignaciones básicas combinadas mediante la fórmula (11). Debido a que el espacio focal es pequeño, las intersecciones que se pueden encontrar son pocas, que son solamente las que se ven en la tabla, obteniendo los valores para  $\varphi_3$  por simple multiplicación.

Tabla III. Cálculo de los potenciales combinados para el diagnóstico de gripa ( $C$ ), resfrío ( $F$ ), migraña ( $H$ ) y meningitis ( $M$ ).

$\varphi_1$	$\varphi_2$	$\varphi_3$
$[\varphi_1(C, F, M)]_m = 0.6$	$[\varphi_2(C, F, H)]_m = 0.7$	$[\varphi_3(C, F)]_m = 0.42$
$[\varphi_1(Q)]_m = 0.4$	$[\varphi_2(C, F, H)]_m = 0.3$	$[\varphi_3(C, F, H)]_m = 0.28$
$[\varphi_1(C, F, M)]_m = 0.6$	$[\varphi_2(Q)]_m = 0.7$	$[\varphi_3(C, F, M)]_m = 0.18$
$[\varphi_1(Q)]_m = 0.4$	$[\varphi_2(Q)]_m = 0.3$	$[\varphi_3(Q)]_m = 0.12$

La marginalización toma un potencial de credibilidad  $\varphi$  en un dominio  $D$  y produce un nuevo potencial  $\varphi^{\downarrow D'}$  en  $D' \subseteq D$ . Es usado para concentrar la información contenida en  $\varphi$  a un dominio más pequeño  $D'$ , y es definida por:

$$[\varphi^{\downarrow D'}(A)]_m := \sum_{B^{\downarrow D'}=A} [\varphi(B)]_m = \sum_{\substack{B^{\downarrow D'}=A \\ B \in FS(\varphi)}} [\varphi(B)]_m \quad (12)$$

En donde  $B^{\downarrow D'}$  denota la proyección del conjunto  $B \subseteq \Theta_D$  al nuevo dominio  $D'$ .

Una nota importante es que la normalización puede ser realizada tanto antes como después de la combinación o marginalización, es decir:

$$v(\varphi_1 \otimes \varphi_2) = v(\varphi_1) \otimes v(\varphi_2) \quad (13)$$

$$v(\varphi^{\downarrow D'}) = v(\varphi)^{\downarrow D'} \quad (14)$$

Así, la normalización puede siempre ser pospuesta hasta el final de los cálculos.

El potencial práctico para la toma de decisiones del marco teórico recién expuesto consiste en:

- 1) Su capacidad para modelar el conocimiento adquirido por el tomador de decisiones a través de observaciones o experimentos, y
- 2) Su habilidad para distinguir entre ignorancia e incertidumbre. Esta última característica es crucial, pues tradicionalmente se ha considerado a la falta de información como una situación indeseable y perjudicial para la toma de decisiones.

Dadas las limitaciones del marco teórico existente hasta hoy, siempre se ha optado por ignorar la incertidumbre, eliminar los datos ambiguos o faltantes y considerar únicamente la información que se conoce con certeza. Sin embargo, DS nos permite adoptar una perspectiva distinta, tratando de aprovechar la falta de información para lograr un mejor conocimiento de la situación del tomador de decisiones.

Para lograr esto, DS considera la ignorancia como la diferencia entre credibilidad y plausibilidad, mientras que al mismo tiempo proporciona una serie de medidas de incertidumbre relacionadas con la entropía,<sup>2, 10, 13, 14</sup> también implementadas en LIDSET pero cuya presentación está más allá del alcance de este artículo.

Así, la teoría de la evidencia enriquece a la probabilidad mediante un marco teórico más general, que permite distinguir entre ignorancia (la diferencia entre la plausibilidad y la credibilidad) e incertidumbre (cuantificada con las medidas de incertidumbre relacionadas con la entropía).

Estos conceptos pueden aportar información valiosa sobre la estructura del problema que sería ignorada usando el enfoque tradicional. Sin embargo, esta teoría tiene algunas desventajas, como su costo computacional para casos no triviales, y la posibilidad de asignar valores positivos al evento vacío, incrementando la incertidumbre sobre el fenómeno.

## COMENTARIOS FINALES

En la vida diaria los seres humanos se enfrentan a situaciones en las que deben tomar decisiones de acuerdo a datos inciertos o incompletos. En este artículo se han mostrado los principales enfoques matemáticos existentes para el tratamiento de la incertidumbre, como la teoría de la probabilidad, las redes bayesianas, la lógica difusa y la teoría de la evidencia de Dempster-Shafer. Se ha explicado por qué se considera que esta última es una de las que tiene mayor potencial y se ha presentado la biblioteca numérica LIDSET para la investigación y desarrollo en el área de razonamiento automático bajo incertidumbre, creada por los autores.

La teoría de la probabilidad es el marco teórico más desarrollado para el manejo de incertidumbre hasta el día de hoy, pero tiene varias limitaciones. Entre ellas están:



- 1) El suponer la existencia de un valor de probabilidad asignado a un evento, independientemente de que el tomador de decisiones tenga la posibilidad de observarlo o no.
- 2) Si este es el caso, se obliga al decisor a estimar estos valores mediante juicios subjetivos.
- 3) Su incapacidad para modelar por separado ignorancia e incertidumbre. Las redes bayesianas resuelven únicamente el primero de estos problemas.

Se ha dicho que el usar valores subjetivos libera al analista de la necesidad de usar la maquinaria teórica de la probabilidad, pues ésta fue desarrollada para el manejo de probabilidades calculadas por conteo; al estimar valores, es tan válido usar probabilidad como otras aproximaciones. Este argumento ha dado lugar al desarrollo de alternativas como la lógica difusa y a la teoría de la evidencia de Dempster-Shafer, cuyos principios básicos se presentaron aquí. Esta última teoría en particular es muy rica en conceptos, adopta de manera natural el punto de vista del tomador de decisiones, y resuelve el problema de modelado de ignorancia e incertidumbre, cuyas magnitudes pueden decir algo sobre la estructura del problema a resolver. Sin embargo, esta aproximación al problema tiene varias desventajas que hoy son materia de investigación, tales como su complejidad computacional. La biblioteca numérica LIDSET es una herramienta con enorme potencial práctico que apoyará el trabajo futuro de investigación.

## REFERENCIAS

1. Luger, G.F., *Artificial intelligence: Structures and strategies for complex problem solving*. Addison Wesley, 4th. Ed., Addison Wesley, Edinburgh, 2002.
2. Beynon, M., Curry, B., Morgan, P., *The Dempster-Shafer theory of evidence: an alternative approach to multicriteria decision modeling*, Omega 28, June 1999.
3. Ross S., *Introduction to Probability Models*, Séptima Edición, Harcourt Academic Press, 2000.
4. Díaz-Ortiz, L., Pacheco-Leal, S., García-Flores, R., *El clasificador Naïve Bayes para la extracción de conocimiento en bases de datos*, Revista Ingenierías, No. 27, 2004.
5. Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, September 1988.
6. Jensen, F. V., *An introduction to Bayesian networks*. Aalborg University. Springer. 1996.
7. Zadeh, L., *Fuzzy sets*, Information and Control, Vol. 8, 338-353, 1965.
8. Sii, H.S., Ruxton, T., Wang, J., *Synthesis using fuzzy set theory and a Dempster-Shafer-based approach to comprise decision-making with multiple-attributes applied to risk control options selection*, Proceedings of IMechE Vol.216 Part E, Pg. 15-29, 2002.
9. Yager, R. R., *Multiple objective decision-making using fuzzy sets*. Int. J. Man-Machine Studies, 1977, 9, 375-382.
10. Klir, G.J., Folger, T., *Fuzzy sets, uncertainty and information*, Prentice-Hall, 1988, London, Toronto, Sydney.
11. Shafer, G., *A mathematical theory of evidence*. Princeton University Press, 1976.
12. Haenni, R., Lehmann, N., *Resource bounded and anytime approximation of belief function computations*, International Journal of Approximate Reasoning 31 (2002) 103–154.
13. Klir, G. J. and Wierman M.J., *Uncertainty-Based Information: Elements of Generalized Information Theory*. Physica-Verlag/Springer-Verlag, Heidelberg and New York, 1999.
14. Beynon, M., Cosker, D., Marshal, D., *An expert system for multi-criteria decision making using Dempster-Shafer theory*. Expert Systems with Applications, 20, 2001, 357-367