

Los Conjuntos Difusos

en la Modelización Lingüística

Francesca Greselin Ferrari

Laboratorio de Electrónica Digital
Universidad Católica Nuestra Señora de la Asunción
Asunción - Paraguay

El análisis de muchos problemas del mundo real necesita el manejo de tipos diferente de informaciones: unos específicos, asertados, y otros vagos, con contornos imprecisos. Los expertos tienen entonces que formular juicios y reglas heurísticas, calificándolos con descriptores lingüísticos como "considerable", "posible", "importante"...

El desarrollo de los Sistemas Expertos necesitó enfocar desde su comienzo el problema de modelar informaciones y relaciones verisímiles aunque no precisas. El ambiente matemático que pudo permitir la introducción de la incertidumbre en la definición de objetos y reglas es la teoría de los conjuntos difusos.

En este artículo se presenta un rápido recorrido en la teoría y una técnica que permite hallar una solución para las operaciones entre conjuntos difusos. Todo esto permite modelar el lenguaje natural a través de variables lingüísticas. Con estas últimas se provee al final una aplicación para previsiones de poblaciones estudiantiles.

INTRODUCCION

El desarrollo de los sistemas expertos empezó su éxito previendo la conducta y dando juicios en campos de la realidad que no pueden ser modelados de manera determinística.

Muchas situaciones reales, en efecto, no permiten partir de **datos ciertos** y/o no pueden ser juzgados con reglas expresadas en **términos matemáticos convencionales**, ya sean determinísticos o probabilísticos.

La información, en general, viene de datos, mediciones o experiencias precedentes. Entonces puede ser vaga, imprecisa, cualitativa, lingüística o incompleta. Los expertos, por lo tanto, además de modelos determinísticos, tienen juicios basados en reglas heurísticas que les permitan lograr previsiones en casos en los cuales hayan factores muy difíciles de medir o cuantificar precisamente. Médicos, psicólogos, ingenieros, manejan informaciones que frecuentemente están acompañadas por calificativos como "serio", "importante", "considerable" y de estos datos realizan evaluaciones a través de procedimientos empíricos, ill-structured.*

La incertidumbre puede ser dividida en tres clases (según lo especificado en el "Workshop on Civil Engineering Applications of Fuzzy Sets", Purdue University, Octubre 1985):

- casualidad natural que causa incertidumbre
- imprecisión que causa difusión **
- ignorancia.

La teoría de las probabilidades maneja matemáticamente el primer tipo de incertidumbre y trata de reducir la difusión al mínimo. Pero, en todo caso, los métodos

* Una regla "ill-structured" es aquella donde la relación entre antecedentes y consecuentes no ha sido suficientemente estudiada o para la cual no hay disponible un modelo matemático razonablemente bueno. En este caso, se pueden usar reglas del tipo IF-THEN para expresar una relación suelta entre antecedentes y consecuentes.

** En este artículo entenderemos 'difusión' como 'fuzziness' y 'difuso' como 'fuzzy'.

probabilísticos están basados en la teoría clásica de los conjuntos, en la cual una proposición o es verdadera o es falsa. Muchas de las proposiciones comunes no son fácilmente clasificables en tal contexto. La teoría de los conjuntos difusos ("fuzzy sets") permite enfrentar y tratar aquellas clases de objetos del mundo real que son '**difusas**' por naturaleza, en el sentido que la **transición entre la pertenencia y la no pertenencia a estas clases es gradual más que brusca**.

De hecho, dado un objeto x y una clase F , la pregunta real en muchos casos no es si x es o no miembro de F , sino **el grado en que x pertenece a F** o, en forma equivalente, el nivel de pertenencia de x a F .

Sea, por ejemplo, U un conjunto de personas:

$$U = \{\text{Alberto, Juana, Ricardo, Manuela, Pablo}\}$$

Si queremos caracterizar el subconjunto A de U que recoja las personas altas de estatura, intuimos que es claro incluir Alberto -que mide 1.92 cm de estatura-, pero es más difícil juzgar en que medida Juana -1.63 cm de estatura-, pertenezca a A .

Cuando los conjuntos difusos son utilizados para modelar informaciones difusas, el análisis implica y necesita de operaciones algebraicas extendidas, es decir, **operaciones entre números difusos**.

En este artículo haremos, entonces, un rápido recorrido en los **conceptos básicos de la teoría de los conjuntos difusos**. En detalle veremos después una técnica basada en la representación de conjuntos difusos a través de α -cortes y de un análisis combinatorio de intervalos. Este método permite hallar una **solución discreta pero exacta para las operaciones entre conjuntos difusos** de una manera muy eficiente y simple.

Presentaremos la **conversión entre variables lingüísticas** (del lenguaje natural) y **variables numéricas** (conjuntos difusos).

Posteriormente, como ejemplo, veremos una **aplicación** para la modelización de estudios en el campo educativo y de paso vamos a presentar también el 'modelo en cascada' para analizar problemas complejos.

En conclusión, se dará también una breve descripción del papel que el procesamiento de informaciones difusas tiene en el campo de los **sistemas expertos**.

BASES DE LA TEORIA DE LOS CONJUNTOS DIFUSOS

Sea U un universo, o sea una colección arbitraria de objetos, conceptos o construcciones matemáticas.

Convencionalmente, si A es un subconjunto de U cuyos elementos son $u_1, \dots, u_n$, entonces A es expresado como:

$$A = \{u_1, \dots, u_n\}.$$

Para nuestros propósitos, sin embargo, es mas conveniente expresar A como la unión de los u_i :

$$A = u_1 + \dots + u_n = \sum_{i=1}^n u_i$$

teniendo en cuenta que, por cada i, j :

$$u_i + u_j = u_j + u_i$$

y:

$$u_i + u_i = u_i.$$

Como una extensión de esta notación, Zadeh (1964) sugirió el concepto de conjunto difuso F de U expresandolo mediante:

$$F = \mu_1 u_1 + \dots + \mu_n u_n$$

o, equivalentemente, como:

$$F = \mu_1/u_1 + \dots + \mu_n/u_n$$

donde μ_i , por $i = 1, \dots, n$, representan el grado de pertenencia de los u_i a F .

A menos que no se establezca lo contrario, los μ_i se asumen con valores en el intervalo $[0, 1]$, con 0 y 1 que denotan 'no pertenencia' y 'plena pertenencia' respectivamente.

El papel que en la teoría de conjuntos clásicos jugaba la función característica (midiendo en forma binaria la pertenencia de cada elemento al conjunto A):

$$\chi_A(x) = \begin{cases} 1 & \text{si } x \text{ pertenece a } A \\ 0 & \text{si } x \text{ no pertenece a } A \end{cases}$$

está ampliado y generalizado a través de la **función de pertenencia** en la teoría de conjuntos difusos, permitiendo describir **niveles diferentes** de pertenencia de cada elemento de F .

La representación de un conjunto difuso finito es una forma lineal en los u_i .

Consistentemente, un subconjunto difuso arbitrario de U puede ser expresado mediante la integral:

$$F = \int_U \mu_F(u) / u$$

donde:

$$\mu_F : U \rightarrow [0, 1]$$

es la función de pertenencia de F , y la integral en U denota la unión de los elementos difusos $\mu_F(u) / u$ en el universo U .

Ejemplo:

Sea X el conjunto de los enteros no negativos, $X = \mathbb{Z}^+$, y sea F el subconjunto difuso de X de los "enteros pequeños". Definimos la función de pertenencia de F como:

$$\mu_{\text{enteros pequeños}}(x) = \{1 + (x/5)^2\}^{-1} \quad x = 0, 1, 2, \dots$$

OPERACIONES ENTRE CONJUNTOS DIFUSOS

Sean F y G dos conjuntos difusos en el universo U .

Sean $\wedge$ y $\vee$ los operadores de mínimo y de máximo respectivamente.

Se define la **unión**:

$$F \cup G = \int_U \{ \mu_F(u) \vee \mu_G(u) \} / u$$

la **intersección**:

$$F \cap G = \int_U \{ \mu_F(u) \wedge \mu_G(u) \} / u$$

el **complemento**:

$$\overline{F} = \int_U \{ 1 - \mu_F(u) \} / u$$

Se muestra fácilmente que valen las identidades:

$$\overline{(F \cup G)} = \overline{F} \cap \overline{G}$$

y

$$\overline{(F \cap G)} = \overline{F} \cup \overline{G}$$

En el mismo contexto, se define que F contiene G si y solo si:

$$F \supset G \iff \mu_F(u) \geq \mu_G(u) \quad \forall u \in U$$

Si $F_1, \dots, F_n$ son subconjuntos difusos de $U_1, \dots, U_n$, entonces el **producto cartesiano** de $F_1, \dots, F_n$ es un subconjunto difuso de $U_1 \times \dots \times U_n$, definido por:

$$F_1 \times \dots \times F_n = \int_{U_1 \times \dots \times U_n} [\mu_{F_1}(u_1) \wedge \dots \wedge \mu_{F_n}(u_n)] / (u_1, \dots, u_n)$$

PRINCIPIO DE EXTENSION

Sea F un conjunto difuso definido en X y sea $y = f(x)$. La imagen de F bajo f se define como el conjunto difuso G definido en Y y dado por:

$$G = \int_U \mu_G(y) / y$$

donde:

$$\mu_G(y) = \mu_F(x) \mid y=f(x)$$

La operación de mapeo puede ser extendida al caso general y esto se conoce como el **principio de extensión** (Zadeh 1973).

Sean $F_1, F_2, \dots, F_n$ conjuntos difusos definidos en $X_1, X_2, \dots, X_n$ respectivamente. Sea f una función:

$$f: X_1 \times X_2 \times \dots \times X_n \longrightarrow Y$$

la imagen difusa G de $F_1, F_2, \dots, F_n$ bajo f tiene como función de pertenencia:

$$\mu_G(y) = \bigvee_{\substack{x_1 \in X_1, \dots, x_n \in X_n \\ y = f(x_1, \dots, x_n)}} [\mu_{F_1}(x_1) \wedge \dots \wedge \mu_{F_n}(x_n)]$$

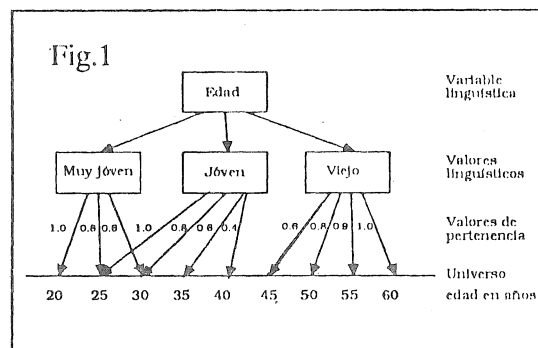
Todas las relaciones matemáticas entre elementos no difusos, y evidentemente también las operaciones algebraicas, pueden ser extendidas a través de este principio en los conjuntos difusos.

Quedan así definidas las **operaciones extendidas**.

EL LENGUAJE NATURAL

La creación de los conjuntos difusos abrió el camino para investigar el lenguaje natural, cuya 'difusión' era conocida desde hace mucho tiempo, permitiendo que este mismo lenguaje pueda ser usado en procedimientos analíticos.

Informalmente, una variable lingüística es una variable cuyos valores -en vez de ser números- son palabras o frases del lenguaje natural (y análogamente se pueden definir para cualquier lenguaje artificial).



Por ejemplo, si "edad" es interpretada como una variable lingüística, entonces el conjunto de sus valores lingüísticos, llamado "term-set", en este caso T (edad), puede ser imaginado como

el que recoge toda expresión del tipo: "jóven", "viejo", "muy jóven", "no muy jóven", "muy muy viejo", "mas o menos jóven", "bastante jóven"...

Cada uno de estos términos es una etiqueta de un conjunto difuso en el universo de las edades, es decir, $U = [0, 100]$.

A una variable lingüística L se le asocian dos reglas:

- a) una regla **sintáctica** que define las sentencias bien estructurada en $T(L)$
- b) una regla **semántica**, a través de la cual el sentido de los términos en $T(L)$ puede ser definido.

Si l es un término en $T(L)$, entonces su significado es un subconjunto de U .

Un **término primario** en $T(L)$ es un término cuyo significado tiene que ser definido a priori, y que sirve como base para computar el sentido de los términos no-primarios en $T(L)$.

Por ejemplo, en $T(\text{edad})$, los términos primarios son "viejo" y "jóven", cuyos significados pueden ser definidos por sus funciones de pertenencia $\mu_{\text{jóven}}$ y μ_{viejo} :

$$\text{"viejo"} = V = \int_U \mu_{\text{viejo}}(u) / u$$

$$\text{"jóven"} = J = \int_U \mu_{\text{jóven}}(u) / u$$

Los términos primarios o atómicos pueden ser conectados entre sí con los conectores lingüísticos "y" (conjunción), "o" (disyunción) -identificables con la operación de intersección y de unión respectivamente- y "no". Además, pueden ser modificados con especificadores lingüísticos como "muy", "ligeramente", "mas o menos"...

Después de haber cuantificado y modelado los términos primarios con los conjuntos difusos, todos los términos compuestos y modificados pueden ser definidos a través de las operaciones extendidas.

Podemos definir:

$$\text{"no jóven"} = J = \int_U [1 - \mu_J(u)] / u$$

$$\text{"muy jóven"} = J^2 = \int_U [\mu_J(u)]^2 / u$$

$$\text{"mas o menos jóven"} = J^{1/2} = \int_U [\mu_J(u)]^{1/2} / u$$

Se puede también hacer el camino al revés, o sea la conversión entre un valor numérico y un término lingüístico, simplemente comparando la función de pertenencia obtenida por medios computacionales y la que más se le acerque entre las de los términos primarios conocidos.

METODO DE LOS VERTICES

El principio de extensión prevee una solución teórica general al algebra de conjuntos difusos. Sin embargo, hallar la solución en muchos casos reales es tarea muy ardua, aún con la ayuda de una computadora.

Bass & Kwakernaak buscaron lograr una solución analítica reconociendo en lo ante dicho un problema de programación no-lineal, pero requiriendo condiciones muy restrictivas y con una implementación bastante engorrosa.

Otros hallaron soluciones aproximadas (Dubois & Prade), (Schmucker, Brown) con métodos de discretización y encontraron problemas a causa de la naturaleza discontinua de las operaciones de máximo y mínimo.

Un nuevo enfoque para evaluar la función de pertenencia fue presentado por Dong y Wong (1985) y se llama **método de los vértices**.

$$y = f(X_1, X_2, \dots, X_n)$$

una función donde los antecedentes $X_1, X_2, \dots, X_n$ son conjuntos difusos y se quiera estimar el consiguiente output conjunto difuso Y .

Para este fin se define el α -corte del conjunto difuso (o conjunto de nivel α de X_j), como:

$$X_j^\alpha = \{x \mid \mu_{X_j}(x) \geq \alpha\} \quad \text{con } \alpha \in [0,1]$$

y también el α -corte de Y como:

$$Y^\alpha = f(X_1^\alpha, X_2^\alpha, \dots, X_n^\alpha)$$

donde los X_j^α son n intervalos que forman una región rectangular n -dimensional con 2^n vértices.

Se puede demostrar fácilmente que si la función f es continua y monótona, entonces:

$$Y^\alpha = [\min_j \{f(c_j)\}, \max_j \{f(c_j)\}] \quad j = 1, \dots, 2^n$$

donde c_j es la coordenada del j -ésimo vértice.

Necesitamos ahora del concepto de modelo en cascada para poder ilustrar después con un ejemplo de aplicación lo que se ha dicho hasta ahora.

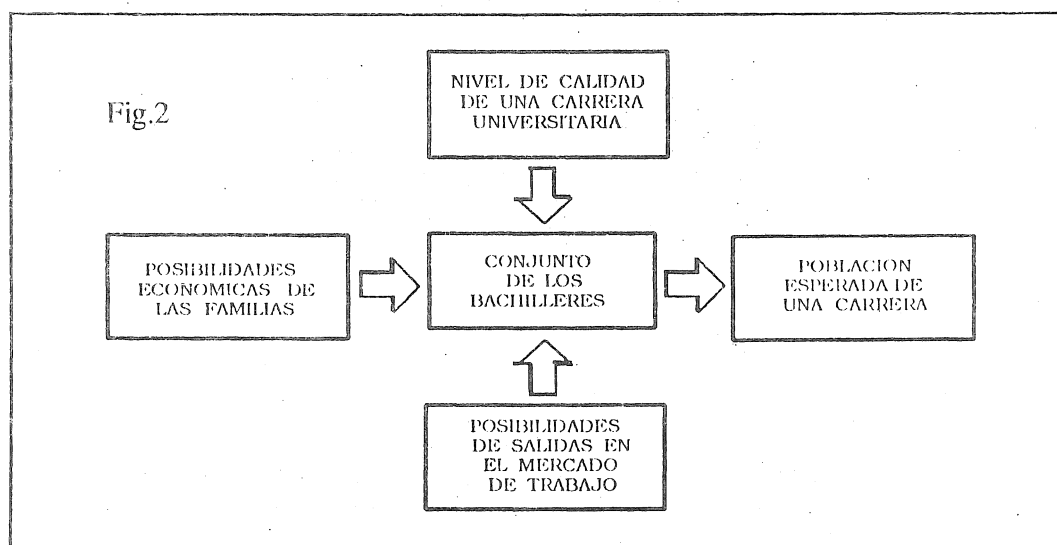
MODELO EN CASCADA

Según el punto de vista bayesiano, hay muchas situaciones reales en las cuales las nuevas informaciones pueden ser utilizadas para actualizar las estimaciones precedentes.

Supongamos que se quiera modelar la expresión lingüística siguiente:

"Si el conjunto de los bachilleres tiene una preparación bastante amplia, y se tiene en cuenta que la posibilidad de las familias para mantener sus hijos en los estudios es medianamente buena, se puede razonablemente esperar una respuesta alentadora en un medio universitario de nivel interesante, considerando que también hayan buenas posibilidades laborales en perspectiva".

Un esquema lógico efectivo podría ser el siguiente:



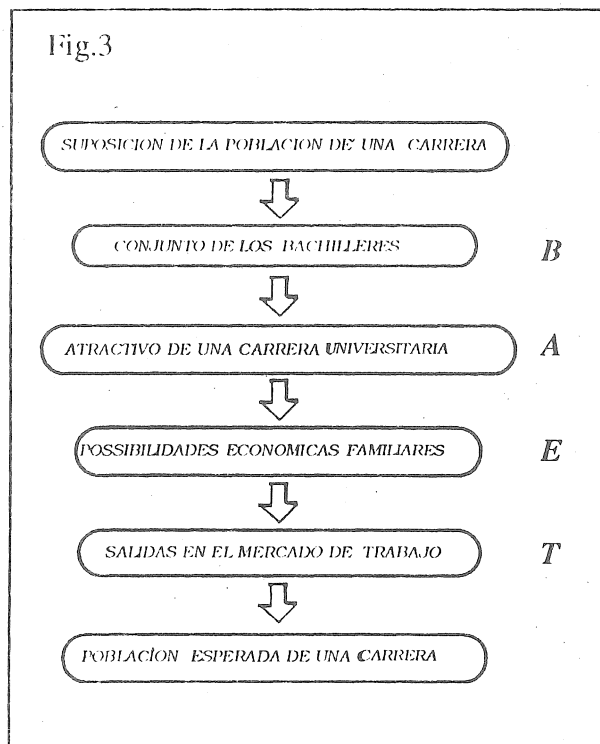
Este esquema resulta complejo porque cada característica que influye en el proceso está independientemente considerada para la evaluación del resultado final, o sea para la evaluación de la población esperable en una carrera.

Para modelar en términos de conjuntos difusos todo este esquema, no tendríamos problemas en describir cada antecedente mediante una función de pertenencia razonablemente adecuada, es decir, mediante una representación de conjuntos difusos. Las "flechas" del dibujo

serían nuestras reglas ill-structured, o sea nuestras reglas difusas. Lo que resulta más difícil de vislumbrar es cómo tener en cuenta **simultáneamente** todos los factores, mediante una regla unificadora.

El procedimiento de evaluación de la población esperada de una carrera puede ser pensado como una **sucesión de decisiones, mientras más y más datos son incorporados en la estimación** - de manera parecida al análisis bayesiano - para actualizar una estimación precedente con nuevas informaciones. Esto permite representar una evaluación compleja a través de una simple serie lineal de operaciones, que parte de una suposición inicial de evaluación.

Aplicémoslo al ejemplo anterior:



Sea entonces X el conjunto difuso que representa la población hipotética de una carrera universitaria y sea Y la evaluación que queremos lograr. Y puede ser obtenido mediante las sucesivas actualizaciones expresadas como:

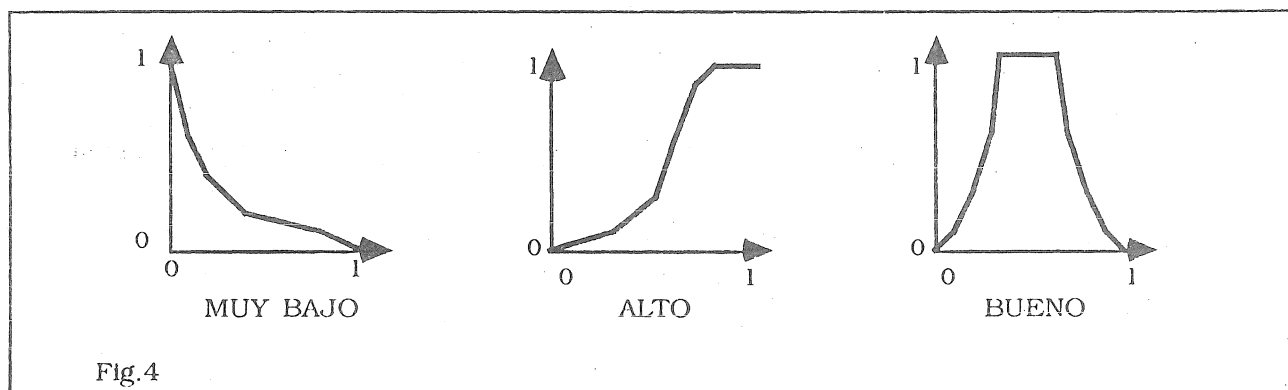
$$Y = \alpha T * \beta E * \chi A * \delta B * X$$

suponiendo que T , E , A , y B sean reglas difusas determinadas por las afirmaciones lingüísticas, que α , β , χ , δ sean pesos a través de los cuales queremos dar diferente importancia a los cuatro factores T , E , A , B asumidos independientes entre ellos, y que '*' sea un operador que indica una relación difusa.

Siguiendo siempre el mismo ejemplo, queremos presentar una aplicación del método de los vértices. Suponemos haber medido que la población X de los bachilleres de una ciudad sea dividida en tres partes iguales, según los diferentes niveles de preparación hallada:

$$\begin{aligned}
 X = & \{ \text{estudiantes con preparación de nivel } \underline{\text{muy bajo}} \} + \\
 & + \{ \text{estudiantes con preparación de nivel } \underline{\text{alto}} \} + \\
 & + \{ \text{estudiantes con preparación de nivel } \underline{\text{bueno}} \}
 \end{aligned}$$

'Muy bajo', 'alto', 'bueno' son términos lingüísticos, cuya representación por conjuntos difusos es:



Juzgamos que sea razonable pensar que frente a la posibilidad de cursar -supongamos- una carrera técnica de nivel interesante, sea muy baja (W_1) la inscripción entre los estudiantes de nivel inferior (X_1), a causa de las pretensiones de la carrera misma. Habrá alta elección (W_2) de la carrera entre estudiantes de buen nivel (X_3) y se supone una media participación (W_3) entre los de alta preparación (X_2) (que pensamos tendrán intereses mejor definidos que los precedentes).

La estimación es entonces una función de las tres componentes de población diferentes por nivel de preparación- pesadas mediante W_1, W_2, W_3 .
Se reconoce que:

$$W_1 = \text{'muy bajo'} = X_1$$

$$W_2 = \text{'alto'} = X_3$$

$$W_3 = \text{'medio'} = X_2$$

Los tres componentes de población X_i , con también los pesos W_i son conjunto difusos ; la población resultante será el conjunto difuso que se halla a través de la media pesada:

$$Y = \frac{\sum_i W_i X_i}{\sum_i W_i}$$

Entonces, de las dos ultimas relaciones obtenemos:

$$Y = f(X_1, X_2, X_3) = \frac{X_1^2 + 2 X_2 X_3}{X_1 + X_2 + X_3}$$

Eligiendo $\alpha = 0,6$ resultan los siguientes α -cortes de los X_i :

$$X_1^{0,6} = [0, 1/10]$$

$$X_2^{0,6} = [3/5, 1]$$

$$X_3^{0,6} = [3/10, 7/10]$$

así que los 2^3 vértices resultantes son:

$$c_1 = (0, 3/5, 3/10)$$

$$c_2 = (1/10, 3/5, 3/10)$$

$$c_3 = (0, 1, 3/10)$$

$$c_4 = (1/10, 1, 3/10)$$

$$c_5 = (0, 3/5, 7/10)$$

$$c_6 = (1/10, 3/5, 7/10)$$

$$c_7 = (0, 1, 7/10)$$

$$c_8 = (1/10, 1, 7/10)$$

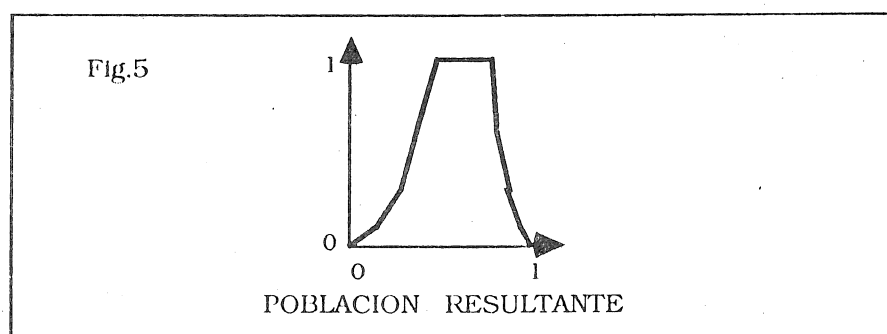
Podemos ahora hallar las $f^{0,6}(c_i)$, obteniendo:

$$f^{0,6}(c_i)_{i=1, \dots, 8} = \{ 2/5, 37/100, 6/13, 61/140, 42/65, 17/28, 14/17, 47/60 \}$$

entonces:

$$Y^{0,6} = [37/100, 14/17]$$

Análogamente, calculando Y^α para diferentes valores de $\alpha \in [0, 1]$, se logra el gráfico de Y:



La población hallada puede entonces ser referida a una población universitaria de 'medio' nivel de preparación (compare este último gráfico con el de X_3 , y haga la conversión de conjunto difuso a variable lingüística).

CONJUNTOS DIFUSOS Y SISTEMAS EXPERTOS

Ya dijimos que gran parte del conocimiento, especialmente el ganado de la observación, es de alguna manera probabilístico y muchas veces también incompleto. Eso significa que se puede considerar algo como verdadero, mientras que aún se permita la posibilidad de un error. Para que una computadora pueda pensar y razonar como un ser humano, se precisa entonces una integración de esta incertidumbre con la lógica.

La mayoría de los hombres manejan la incertidumbre que viene de la observación natural sin muchos problemas, porque están acostumbrados a ella. Sin embargo, hacer posible que una computadora sepa trabajar con la incertidumbre, necesita de un buen esfuerzo. En el lenguaje de los programadores, es una tarea no banal la de juntar la lógica con la incertidumbre, porque son opuestas.

La lógica difusa nos permite pensar en valores verdaderos en el sentido, más débil pero mucho más próximo a la realidad, de ser verosímiles.

Es fácil intuir como, para los problemas de reconocimiento y clasificación de patrones (pattern recognition y pattern classification) y para los sistemas expertos, el manejo de la incertidumbre vuelve a ser un punto crítico de evaluación de las posibilidades de la Inteligencia Artificial.

Se puede subdividir el problema en dos pasos distintos:

- a) convertir el dato p en un objeto matemático $M(p) = \{M_1(p), \dots, M_n(p)\}$
- b) pasar de algoritmos 'opacos' A_{op} a algoritmos 'transparentes' A_{tp} , o sea de algoritmos cuyas descripción no está todavía conocida, a algoritmos que actúen sobre el conjunto $M(p)$ realizando el grado de pertenencia de p como definido por A_{op} .

Es importante observar que desde un punto de vista práctico, es deseable que:

- a) $M(p)$ sea definido por un pequeño número de atributos
- b) la medición de estos atributos sea relativamente simple.

En la casi totalidad de los Sistemas Expertos, si se desea seguir la formulación descripta, se encuentran problemas para lograr una buena definición matemática clásica.

Ya vimos que los métodos de la lógica difusa se pueden aplicar ya sea en la evaluación de los datos, como en la construcción de reglas 'sueltas'. La estructura conceptual de la teoría de los conjuntos difusos provee entonces un ambiente más natural para la formulación y la solución aproximada de estos tipos de problemas, mucho más apta que los métodos tradicionales, basados en la teoría clásica de los conjuntos, en la teoría de las probabilidades y en la lógica booleana.

Bibliografía

- Baas S. and Kwakernaak H. (1979), "Rating and Ranking of Multiple Aspects Alternatives Using Fuzzy Sets". Automatica, 13.
- Brown C.B. and Yao J.T.P. (1983), "Fuzzy Sets and Structural Engineering". ASCE J. Struct. Eng., 109, ST5.
- Dong W.M. and Shah H.C. (1985), "Vertex Method for Computing Function of Fuzzy Variables", Journal of Fuzzy Sets and Systems.
- Dubois D. and Prade H. (1980), "Fuzzy Sets and Systems: Theory and Applications". Academic Press, New York.
- Duda R., Hart P. and Nilsson N. (1976), "Subjective Bayesian Methods for Rule-based Inference Systems". Proceedings of National Computer Conference, pp. 1075-1082.

Kaufmann A. and Gupta M.M. (1985), *"Introduction to Fuzzy Arithmetic: Theory and Applications"*. Van Nostrand Reinhold Company, New York.

Schmucker K.J. (1984), *"Fuzzy Sets, Natural Language Computational, and Risk Analysis"*. Computer Science Press, Potomac, Maryland.

Wong F.S., Dong W., Bolssonade A. and Ross T.J. (1986) *"Expert Opinions and Expert Systems"*. ASCE 9th Electronic Computations Conf., Birmingham, Alabama.

Zadeh L.A. (1964), *"Fuzzy Sets"*, J. Fuzzy Sets and Systems, Memo, EREL., No. 64-44, University of California, Berkeley.