

# Modelo difuso para la selección automatizada cursos apropiados para el perfil profesional del alumno

Corredera de Colso, Luis E.<sup>1</sup>; López Rivero, Alfonso<sup>2</sup>; Vallejo García, Marcelo<sup>2</sup>

1. Flag Solutions S.L. Salamanca, España.

Email:luisenrique@flagsolutions.net. WWW-HOME: www.flagsolutions.net

2. Universidad Pontificia de Salamanca. Salamanca, España.

Email:ajlopezri@upsa.es, mvallejoga@upsa.es

**Abstract.** En este trabajo presentamos un modelo matemático inspirado en conjuntos difusos para la ayuda en la selección y clasificación de los profesores y estudiantes conforme a su base de conocimientos reglados. Este modelo hace uso de la teoría de conjuntos difusos y la proyección geométrica para conseguir un método computacionalmente eficiente en su objetivo, que es capaz de clasificar los individuos en niveles de adecuación, tratando de forma conveniente tanto los excesos como los defectos en la adecuación de los candidatos.

## Introducción a la selección de personal

El problema de la selección automatizada de personal para cubrir determinados puestos de trabajo de forma óptima en una organización es un problema ampliamente considerado y estudiado [1,2], y cuyo éxito determina el cumplimiento de los objetivos y proyectos, especialmente en pequeñas empresas [5]. Entre las diferentes propuestas publicadas, existen soluciones que se apoyan fundamentalmente en la teoría de conjuntos difusos, en ocasiones combinadas con algoritmos genéticos y evolutivos [2,3,4].

El problema de la selección automatizada de personal en las empresas es fácilmente extrapolable a otros terrenos prácticos de aplicación, como la gestión de la formación en la carrera profesional de los empleados. En nuestro caso concreto, la aplicación es conveniente para optimizar dos aspectos de la formación online:

- Poder ofrecer cursos adaptados a la formación de base de los alumnos.
- Poder seleccionar personal docente que tenga cualificación y conocimiento de base suficiente para atender la formación de sus alumnos con solvencia.

## Ventajas y limitaciones de los métodos difusos y evolutivos

Como se ha citado anteriormente, existen múltiples enfoques para la resolución de este problema, principalmente basados en lógica de conjuntos difusos [6,7] y algoritmos genéticos [4,8].

La ventaja que el empleo de conjuntos difusos aporta sobre el problema es la capacidad para clasificar la pertenencia de elementos a diferentes conjuntos, con una función de pertenencia cuyo valor puede ser un número real.

El enfoque de aplicación de algoritmos evolutivos al problema proporciona una mejora continua, a través del mezclado poblacional y extracción de las mejores características, sin embargo su complejidad y coste computacional es muy superior al enfoque propuesto por nuestro trabajo.

## Descripción de nuestro sistema difuso

Nuestro trabajo para la obtención del modelo difuso [9] presentado se llevó a cabo a través de:

La definición de las condiciones iniciales y premisas de funcionamiento, el preprocesamiento de las variables de entrada y salida del modelo.

Estimación de parámetros del sistema, de modo que se comporte de acuerdo con las condiciones iniciales.

La verificación del modelo, y la cuantificación de su error.

La validación del modelo: asegurando que soluciona los problemas planteados, y se comporta de acuerdo con su especificación.

Al enfrentarnos inicialmente al problema hemos considerado la conveniencia de establecer unas condiciones iniciales y unas premisas de funcionamiento que han de cumplirse en los resultados del modelo:

1. Existe la necesidad de codificar las variables cualitativas con las que vamos a trabajar. En nuestro caso, las titulaciones de las personas que serán seleccionables.
2. Se hace notar la conveniencia de establecer las distancias existentes entre los diferentes pares de elementos de las variables codificadas anteriormente. En este caso, la distancia no es una relación conmutativa: la distancia entre A y B no tiene por qué ser igual a la distancia entre B y A. Esta distancia podemos formularla en lenguaje natural mediante la siguiente fórmula “si

- necesito un titulado en A, cómo de válido es una persona con la titulación B”. Lógicamente el operador de distancia no será conmutativo.
3. Existe la posibilidad de que una persona tenga dos o más titulaciones que sean cercanas (incluso alguna coincida exactamente) según la concepción de distancia establecida anteriormente, en cuyo caso éstas deben contribuir a una mejora del resultado. En caso de que una de ellas coincida con el objetivo y exista alguna titulación cuya distancia sea “finita”, ésta debe de contar para poder valorar la ordenación también de las características excedidas.
  4. Existe la posibilidad de que una persona tenga una titulación conveniente y otra cuya distancia sea infinita (en nuestro lenguaje). En cuyo caso, la titulación de distancia infinita no debe de contribuir a una peor evaluación del perfil profesional (por ejemplo, un ingeniero técnico en informática de sistemas y un diplomado en empresariales son titulaciones complementarias, pero si el objetivo de búsqueda es un diplomado en empresariales, la ingeniería técnica no debería aumentar la distancia al no tener en principio ninguna relación). El conocimiento en este caso no es suplementario, sino complementario y como tal ha de ser modelado.

### **Codificación de las variables cualitativas y establecimiento de distancias**

Los valores numéricos asignados para las distancias determinan el éxito del funcionamiento del algoritmo. Se establecieron los siguientes criterios para valorar la distancia entre dos titulaciones (tomadas en pares objetivo-existente):

- Cuando la una titulación no tenga absolutamente nada que ver con otra se le asigna un valor “1”. Al definir en el algoritmo la distancia entre una persona y el objetivo como el producto de las distancias de todas sus titulaciones, el “1” como elemento neutro del producto nos ayuda al cumplimiento de la cuarta premisa establecida previamente.
- Cuando la titulación es idéntica o equivalente, se asigna un valor de “-1” a la distancia. De esta forma, habiendo definido la distancia de una persona al objetivo como el producto de todas sus distancias, la proyectamos directamente sobre  $R^-$ , indicando a nuestro algoritmo que esa persona es al menos un match que cumple el criterio establecido.
- Cuando la titulación objetivo tiene alguna afinidad distinta de idéntica, el valor de su distancia será un número real perteneciente al intervalo abierto  $(0,1)$ . Este criterio se establece numéricamente completando la afirmación formulada en lenguaje natural como “cuando necesito un titulado en A y obtengo un titulado en B, es válido en un X %”. Al valor X, expresado en tantos por 1, y sustraído del elemento neutro es el valor numérico de la distancia.

La concepción de distancia para nuestro modelo transgrede las medidas de comparación de distancia basada en las distancias de Minkowski y sus casos particulares (Hamming, euclídea, Tchebyshev) en que introduce el concepto de distancia “negativa”, por conveniencia para el tratamiento del modelo.

Los valores para las distancias son datos de entrada al sistema, que deberían ser recogidos por la opinión y experiencia de expertos en la materia.

Para la puesta en producción del modelo en los casos referidos al final del artículo se codificó cada titulación universitaria reconocida en España con un identificador único numérico. Los valores de las distancias se calcularon a través de la realización de una encuesta distribuida a profesores de universidades españolas, que posteriormente fueron tratados estadísticamente para eliminar los sesgos en las valoraciones e incorporarlos al modelo como los valores iniciales del sistema (matriz de distancias).

### **Cálculo de la adecuación de una persona a un objetivo**

Para la obtención del cálculo de adecuación de una persona a un objetivo, el algoritmo es el siguiente:

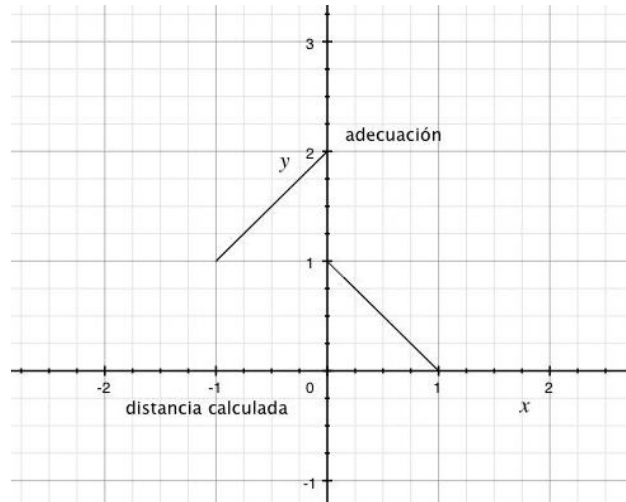
- 1.- Establecemos la adecuación inicial en “1”.
- 2.- Para cada titulación de la persona, multiplicamos la adecuación por la distancia.

$$a_0 = 1;$$
$$a(p, o) = a_0 \cdot \prod_p d(t_p, o)$$

### **Ecuación 1: Cálculo de la adecuación de una persona a los objetivos**

Los resultados de este algoritmo nos permiten posicionar a todos los candidatos en una recta, en el intervalo cerrado  $[-1,1]$ , siendo su interpretación la siguiente según el valor “a” de su función de adecuación:

- a pertenece a  $[-1,0)$ : el perfil del candidato cumple el objetivo, y tiene un exceso. Este exceso se obtiene cuantitativamente como el valor absoluto de “a”, siendo éste mayor, indica un mayor exceso de requisitos.
- a pertenece a  $[0,1]$ : el perfil candidato no cumple el objetivo, sin embargo cuanto más cercano a cero sea, mayor será su grado de adecuación al objetivo.



**Figura 1: Interpretación de los valores de salida de la función de pertenencia**

### Proyección geométrica de la salida del algoritmo

La función de salida del algoritmo entrega los resultados en un orden que no resulta óptimo para su tratamiento en cantidades masivas de forma eficiente, por ejemplo para la aplicación de algoritmos de ordenación.

Por ello a la salida del algoritmo se le ha aplicado una función de transformación, que proyecta la entrada de forma ordenada sobre la parte positiva de la recta de los números reales en el intervalo  $[0,2]$ :

La función de proyección es la siguiente:

$$y = f(x) \begin{cases} [-1, 0) : x + 2 \\ [0, 1] : 1 - x \end{cases}$$

### Ecuación 2: Proyección de la salida para ordenarla

Esta función establece por tanto dos alpha-cortes en el conjunto de salida, proporcionando dos conjuntos difusos y un conjunto clásico para su salida, con las siguientes funciones de pertenencia:

$$\mu(y) : \begin{cases} \text{Inferiores} : \mu(y) \in [0, 1[ \\ \text{Exactos} : \mu(y) = 1 \\ \text{Excesivos} : \mu(y) \in ]1, 2[ \end{cases}$$

**Ecuación 3: Intervalos definidos por los alpha-cortes**

De esta forma, los valores de adecuación quedan ordenados naturalmente por su valor numérico y se facilita tanto la comparación entre pares como su almacenamiento y posterior extracción de forma ordenada en sistemas gestores de bases de datos.

**Ejemplos sintéticos de funcionamiento**

Durante la fase de análisis del dominio del problema, diseño del algoritmo, validación con respecto de las restricciones iniciales y optimización definimos de forma sintética las distancias entre pares de objetivos y potenciales como muestra la tabla.

		Perfiles potenciales encontrados			
		ISI	IST	ITIS	DUE
Objetivos buscados	ISI	-1	0,7	0,5	1
	IST	0,7	-1	0,8	1
	ITIS	0,4	0,6	-1	1
	DUE	1	1	1	-1

**Tabla 1:Distancias entre pares**

La interpretación de los datos debe hacerse teniendo en cuenta que cuando el perfil potencial presenta una coincidencia exacta, recibe el valor “-1” que nos ayuda a proyectar su adecuación sobre otra región del espacio geométrico que será decodificada posteriormente.

Cuando las titulaciones no son afines, son valorados con el elemento neutro de la multiplicación, para evitar la distorsión que pueda introducir en casos de existencia de varias titulaciones en el potencial.

Cuando las titulaciones son afines, presentan una distancia que significa mayor adecuación cuando está más cercana a cero. De modo, que lingüísticamente podríamos interpretarlo como “cuando busco un IST, y encuentro un ITIS, éste es válido en un 30%” (el valor 30% se obtiene de sustraer el coeficiente presente en la matriz de la unidad, y expresarlo en tantos por cien).

Veamos a continuación su comportamiento en diferentes problemas sintéticos.

### Comportamiento ante búsqueda con coincidencia exacta

Enunciamos el caso de prueba sintético como “cuando el objetivo es un titulado en ISI, ¿cómo de válidos son nuestros potenciales?”

La siguiente tabla muestra los resultados de la evaluación, tanto el resultado intermedio como el resultado final:

Cuando busco	Y obtengo	Adecuación	salida final
ISI	ISI	-1	1
ISI	IST	0,7	0,3
ISI	ITIS	0,5	0,5
ISI	DUE	1	0

**Tabla 2: Ejemplo de problema sintético con resultado exacto**

Los resultados denotan una salida exacta, y una afinidad mayor entre la titulación ITIS y la titulación ISI que entre la titulación IST y la titulación ISI. Queda fuera de la adecuación, totalmente, una titulación DUE cuando el objetivo planteado era una titulación ISI.

### Una búsqueda con resultados excesivos, exactos y complementarios

Enunciamos el caso de prueba sintético como “cuando necesito un titulado en IST, ¿cómo de válidos son nuestros potenciales?”

La siguiente tabla muestra los resultados de la evaluación, tanto el resultado intermedio como el resultado final:

Cuando busco	Y obtengo	Adecuación	salida final
IST	ISI + IST	-0,7	1,3
IST	IST	-1	1
IST	ISI	0,7	0,3
IST	ISI+DUE	0,7	0,3

**Tabla 3: Resultados para búsqueda con niveles excesivos, exactos e inferiores**

Los resultados denotan que la obtención de un potencial con dos titulaciones, una exacta y una complementaria tendría una adecuación (1,3) por encima de la exacta (1).

También se advierte en el resultado que el hecho de que aparezca una titulación inadecuada (DUE) para el objetivo en conjunto con una titulación parcialmente adecuada no penaliza al caso en que solamente se presenta la titulación parcialmente adecuada (resultados 3 y 4 de la tabla).

**Una búsqueda excesiva, con diferentes niveles de exceso.**

Enunciamos el caso de prueba sintético como “cuando necesito un titulado en IST, ¿cómo de válidos son nuestros potenciales?”

La siguiente tabla muestra los resultados de la evaluación, tanto el resultado intermedio como el resultado final:

Cuando busco	Y obtengo	Adecuación	salida final
ITIS	ITIS+HSI	-0,4	1,6
ITIS	ITIS+HST	-0,6	1,4
ITIS	ITIS	-1	1
ITIS	ITIS+DUE	-1	1

**Tabla 4: Cuantificación de los excesos**

La interpretación de los resultados denota, la existencia de perfiles excesivos debido a la acumulación de titulaciones, así como muestra nuevamente que las titulaciones que no están relacionadas con los objetivos, se comportan de forma neutra con los resultados.

**Ejemplos reales de funcionamiento**

El modelo difuso aquí propuesto ha sido particularmente empleados, con sus adaptaciones, a diferentes aplicaciones para entornos de producción. Su implementación ha sido realizada por la empresa Flag Solutions S.L. en proyectos de ámbitos tan dispares como herramientas para la selección de recursos humanos de forma automatizada, como en un portal B2B de Internet para la contratación de maquinaria industrial para proyectos de obra civil, funcionando con un desempeño óptimo.

Para la optimización de su rendimiento y conseguir una experiencia de usuario mejorada incluso en situaciones de alta concurrencia se diseñó la realización del precálculo de las distancias entre todos los potenciales existentes y los objetivos existentes con cada inserción de un potencial, y con cada inserción de un objetivo, en

un proceso por lotes desatendido. La información de distancia entre pares se almacena en base de datos normalizada, permitiendo su uso de forma inmediata.

## **Líneas de trabajo futuro**

Dada la amplia aplicabilidad del modelo para diferentes ámbitos del mundo real, y la selección y ordenación numérica atendiendo a valoraciones definidas en ámbitos pertenecientes a las ciencias sociales, serían líneas de trabajo futuro para la mejora del modelo las siguientes:

- Modelado de pares de distancia para diferentes situaciones del mundo real (experiencia, ámbitos de conocimiento, motivación, etc.) y su forma de asignarles una valoración coherente con el modelo presentado.
- Extensión del modelo para criterios multi-variable, facilitando en la definición de objetivos establecer la importancia que representa cada variable estudiada.
- Mejorar el modelo para incluir factores de corrección procedentes de realimentación por parte del usuario, de forma que puedan modificarse particularmente los resultados de clasificación, en contra de los datos de configuración inicial.
- En otro ámbito, estudio del impacto de los excesos en factores psicológicos para recursos humanos (frustración en recursos humanos con un exceso considerable, incremento de motivación por asunción de responsabilidades ligeramente superiores a la capacitación presentada, etc.).

## **Conclusiones**

La selección de recursos es un proceso determinante para el éxito tanto de los proyectos como de las empresas.

El modelo expuesto en este trabajo ha demostrado a través de los problemas sintéticos y de su funcionamiento en entornos de producción la solvencia para obtener conjuntos de recursos apropiados para los objetivos, ordenables de forma numérica y con la posibilidad de conocer excesos y defectos.

## **Referencias**

- [1] Canós, L; Liern, V. (2004). Some Fuzzy Models for Human Resources Management. *International Journal of Technology, Policy and Management*, Vol.4, No.4, pp.291-308.
- [2] Canós, L.; Caño, C; González, B (2006). Algunos algoritmos de ordenación para el proceso de selección de personal. X Congreso de Ingeniería de Organización.

- [3] Cañós, L.; Cassaús, T.; Lara, T; Liern, V; Pérez, J. (2007) Un algoritmo fuzzy para la selección de personas basado en competencias. XV Jornadas de ASEPUMA y III Encuentro Internacional.
- [4] Osorio, L.; Fernández, E. Merlino, H; Rodríguez, D; Britos, P; García Martínez, R; (2007) Utilización de algoritmos evolucionarios para la asignación de recursos. III Congreso argentino de ciencias de la computación.
- [5] Corredera, L.E.; López, A.; Vallejo, M; Joyanes, L.;(2007) Costes de omisión de prácticas de calidad en la pequeña empresa. 6th International Workshop on Practical Applications of Agents and Multiagent Systems.
- [6] Zadeh, L. (1965): "Fuzzy Sets". Information and Control, Vol. 8, pp. 338-357.
- [7] Kruse, R; Gebhardt, J; Klawonn, F; Foundations of Fuzzy Systems. John Wiley & Sons, 1994.
- [8] Goldeberg, D. E. (1989). "Genetic Algorithms in Search Optimization & machine learning". Addison-Wesley, Massachusetts.
- [9] Bezdek, J.C. (1993). Fuzzy Models: What are They, and Why?. IEEE Trans. On Fuzzy Systems 1, pp. 1-6.