

Aprendiendo consultas lingüísticas difusas con algoritmos evolutivos multiobjetivo. Estudio preliminar⁰

A.G. Lopez-Herrera F. Herrera, E. Herrera-Viedma, A. Zafra
Dpto. de Informática, Dpto. Ciencias de la
Universidad de Jaén Computación e I.A.,
aglopez@ujaen.es Universidad de Granada
{herrera, viedma}@decsai.ugr.es

M. Luque
Dpto. Ciencias de la
Computación y A.N.,
Universidad de Córdoba
mluque@uco.es

Resumen

El rendimiento de un sistema de recuperación de información es generalmente medido en términos de dos criterios diferentes, precisión y exhaustividad. De este modo, la optimización de cualquiera de sus componentes es un claro ejemplo de problema multiobjetivo. En esta contribución abordaremos el problema del aprendizaje automático de consultas mediante técnicas evolutivas multiobjetivo, en concreto nos centramos en el aprendizaje de consultas lingüísticas difusas. Presentamos un estudio comparativo de tres de las técnicas de optimización multiobjetivo de propósito general más conocidas, NSGA-II, SPEA2 y MOGLS, en el contexto de los sistemas de recuperación de información difusos.

Palabras Clave: Sistemas de Recuperación de Información, Programación Genética, Inductive Query By Example, Algoritmos Evolutivos Multiobjetivo, Aprendizaje de Consultas.

1 INTRODUCCIÓN

La Recuperación de Información (RI) puede ser definida, en general, como el problema de la selección de información documental desde una fuente de almacenamiento en respuesta a necesidades de información del usuario, expresadas éstas mediante consultas [1, 21]. Los Sistemas de Recuperación de Información (SRI) son un tipo de sistemas automáticos

de información que trabajan con bases de datos compuestas por elementos de información, documentos que pueden ser textuales, gráficos o registros sonoros, y procesan las consultas de los usuarios permitiéndoles acceder a la información relevante en un intervalo de tiempo adecuado.

La mayoría de los SRI basados en el modelo Booleano [25] presentan algunas limitaciones. Debido a ello, se han diseñado otros paradigmas con el fin de extender este modelo de recuperación y superar sus problemas, como los modelos de RI difusa [5, 19, 24].

Sin embargo, el aumento en la potencia del modelo de recuperación también conlleva un aumento en la complejidad del lenguaje de consulta, tanto en el modelo Booleano como en el difuso. Por esto, se hace necesario diseñar métodos automáticos para desarrollar esta tarea.

De este modo, el paradigma "Inductive Query by Example (IQBE)" [6], consistente en el aprendizaje automático de una consulta a partir de un conjunto de documentos proporcionados por un usuario, puede ser útil para resolver este problema y asistir al usuario en el proceso de formulación de consultas.

Los Algoritmos Evolutivos (EAs) se han empleado para el aprendizaje automático de consultas [8]. Estos algoritmos se caracterizan por tener un conjunto de soluciones (consultas en nuestro contexto) de prueba a las que les aplican cambios aleatorios, y mediante mecanismos de selección determinan que soluciones serán mantenidas en futuras generaciones. En particular, las técnicas de Programación Genética (PG) [17], donde las consultas son representadas como árboles sintácticos de expresión y adaptados mediante procesos de selección, cruce y mutación, han sido las técnicas de EAs más utilizadas.

Dos de los primeras aproximaciones del IQBE al ámbito del aprendizaje automático de consultas son las propuestas por Smith y Smith [23] en el contexto

⁰Este trabajo ha sido financiado conjuntamente por el proyecto GLIRS-II, Ref. TIC-2003-07977 y el proyecto de excelencia de la Junta de Andalucía SAINFOWEB, Cod. 00602.

del modelo Booleano, y la de Kraft y otros [18] en el modelo difuso. Ambos guiados por una función fitness que combina, precisión y exhaustividad en un mismo objetivo.

En este trabajo abordaremos un estudio comparativo del desempeño, en el aprendizaje automático de consultas difusas con ponderaciones lingüísticas, de tres de los Algoritmos Evolutivos Multiobjetivo (MOEAs) más utilizados recientemente en la literatura, como son NSGA-II [10], SPEA2 [27] y un MOEA combinado con local search MOGLS [12], adaptados a la estructura de PG con árboles sintácticos de expresión codificando las consultas. Todos ellos optimizando, de manera simultánea, ambos objetivos, precisión y exhaustividad.

Para ello estructuraremos la contribución como sigue. La Sección 2 se dedicará a introducir las bases de los SRI difusos, así como la técnica IQBE. La descripción y particularización de los MOEAs utilizados se llevará a cabo en la Sección 3. En la Sección 4 se expondrá la experimentación, y finalmente, en la Sección 5 se recogerán algunas conclusiones.

2 PRELIMINARES

En esta sección introduciremos las bases de los SRIDs, sus componentes y procedimiento de evaluación, así como una breve descripción de la técnica IQBE.

2.1 Sistemas de Recuperación de Información Difusos

Los Sistemas de Recuperación de Información Difusos (SRIDs) permiten manejar la incertidumbre y la imprecisión inherentes al proceso de recuperación extendiendo los SRI Booleanos. Sus tres componentes son [2]:

Base de datos: almacena los documentos y su representación (típicamente basada en términos índice en el caso de documentos textuales). Sea \mathcal{D} un conjunto de documentos y \mathcal{T} un conjunto de términos representando el contenido subyacente de estos documentos. Existe una función de indización $\mathcal{F} : \mathcal{D} \times \mathcal{T} \rightarrow [0, 1]$ que define un conjunto difuso bidimensional (una relación difusa), que se proyecta para obtener el conjunto difuso asociado a cada documento y términos individuales:

$$d_i = \{(t, \mu_{d_i}(t)) | t \in \mathcal{T}; \mu_{d_i}(t) = \mathcal{F}(d_i, t)\}$$

$$t_j = \{(d, \mu_{t_j}(d)) | d \in \mathcal{D}; \mu_{t_j}(d) = \mathcal{F}(d, t_j)\}$$

Existen diferentes formas para definir la función de indización \mathcal{F} , la más extendida es la frecuencia inversa del documento (IDF) [21].

Subsistema de consulta: permite a los usuarios formalizar sus consultas. Las consultas difusas son expresadas usando un lenguaje de consulta basado en la ponderación de términos como en [3], donde los pesos numéricos o lingüísticos están asociados a alguna semántica (importancia relativa, cuantitativa o umbral) modelando algún requerimiento del usuario y los términos son conectados mediante conectivos lógicos *AND* y *OR*.

En SRIDs el subsistema de consulta proporciona un conjunto difuso q definido sobre el dominio de los documentos especificando el grado de relevancia (el así llamado *retrieval status value* (RSV)) de cada documento en la base de datos con respecto a la consulta procesada:

$$q = \{(d, \mu_q(d)) | d \in \mathcal{D}; \mu_q(d) = RSV_q(d)\}.$$

Subsistema de evaluación: evalúa el grado para el cual la representación del documento satisface los requerimientos expresados en la consulta (es decir, el RSV) y recupera aquellos documentos que son juzgados como relevantes a ella. Cuando usamos la semántica de umbral, los pesos de la consulta representan requerimientos mínimos que la representación de los documentos deben satisfacer para que estos sean considerados relevantes. La evaluación de una consulta se realiza mediante un proceso constructivo ascendente, consistente en evaluar en primer lugar los términos y en segundo lugar las combinaciones de estos aplicando los conectivos lógicos *AND* y *OR*.

2.2 Evaluación de los Sistemas de Recuperación de Información

Existen diferentes maneras de medir la calidad de un SRI, tales como la eficiencia y la eficacia, y varios aspectos subjetivos relacionados con la satisfacción del usuario [1]. Tradicionalmente, la eficacia de la recuperación - basada en la relevancia de los documentos con respecto a las necesidades de información del usuario - es la más empleada. Hay diferentes criterios para medir este aspecto, siendo la precisión P y la exhaustividad E las más usadas. La precisión se define como la razón entre el número de documentos relevantes recuperados por el SRI en respuesta a una consulta y el número total de documentos recuperados, mientras que la exhaustividad es la razón entre el número de documentos relevantes recuperados y el número total de documentos relevantes existentes en la base de datos [25]. La expresión matemática para cada una de ellas es:

$$P = \frac{D_{rr}}{D_{tr}} \quad ; \quad E = \frac{D_{rr}}{D_{rt}}$$

nite a los usuarios las consultas difusas en lenguaje de consulta términos como en [3], lingüísticos están aso- importancia relativa, dando algún requer- inos son conectados AND y OR.

consulta proporciona sobre el dominio de el grado de relevan- tus value (RSV)) de : datos con respecto

$$= RSV_q(d).$$

lúa el grado para el imento satisface los la consulta (es de- os documentos que a ella. Cuando us- os pesos de la con- os mínimos que la os deben satisfacer los relevantes. La : realiza mediante términos y en se- de estos aplicando DR.

s de ión

la calidad de un acia, y varios as- a satisfacción del cacia de la recu- de los documen- : información del diferentes crite- la precisión P y La precisión se o de documentos respuesta a una entos recupera- s la razón entre recuperados y el es existentes en matemática para

donde D_{rr} es el número de documentos relevantes recuperados, D_{lr} es el número total de documentos recuperados y D_{rt} es el número total de documentos relevantes existentes en la base de datos.

Ha de tenerse en cuenta, que ambas medidas están definidas en el intervalo $[0,1]$, siendo 1 el valor óptimo. Obsérvese que el único medio de conocer todos los documentos relevantes existentes para una consulta en una base documental (valor necesario para calcular la exhaustividad) es evaluarlos todos uno por uno. Debido a esto y a que la relevancia es subjetiva, hay disponibles diversas bases documentales clásicas, cada una con un conjunto de consultas para las que se conocen los juicios de relevancia, de forma que puedan ser usadas para verificar las nuevas propuestas que se dan en el campo de la RI [1, 20]. En esta contribución, utilizaremos la colección Cranfield.

2.3 La técnica IQBE

La técnica IQBE fue propuesta por Chen [6] como un proceso en el que los usuarios proporcionan documentos (ejemplos) y el algoritmo induce (o aprende) los conceptos claves con el fin de encontrar otros documentos similares (e igualmente relevantes). De esta forma, IQBE podría entenderse como una técnica para asistir a los usuarios en el proceso de formulación de consultas desarrollado por medio de métodos de aprendizaje automático. Trabaja tomando un conjunto de documentos relevantes (y opcionalmente documentos no relevantes) proporcionados por el usuario - que pueden ser obtenidos de una consulta preliminar o de un proceso de hojear de la base documental - y aplicando un proceso de aprendizaje para generar automáticamente una consulta que describa las necesidades de información del usuario (representadas por los documentos proporcionados por él). La consulta que se obtiene puede ejecutarse en otros SRIs para obtener nuevos documentos relevantes. De esta forma, no hay necesidad de que el usuario interactúe con el proceso como ocurre en otras técnicas de refinamiento de consultas tales como la retroalimentación por relevancia [20].

Se han propuesto métodos IQBE para los diferentes modelos de RI existentes [8]. Los métodos IQBE más utilizados son los basados en PG, donde las consultas son representadas mediante árboles sintácticos de expresión y los algoritmos son articulados en base a los operadores clásicos: cruce, mutación y selección.

2.4 Modelo Lingüístico Difuso Ordinal

Un enfoque lingüístico difuso ordinal se define considerando un conjunto de etiquetas finito y totalmente ordenado $S = \{s_i, i \in \{0, \dots, T\}\}$ con $s_i \geq s_j$ si $i \geq j$, y con una cardinalidad impar (la cardinalidad de S

es $T + 1$). La semántica del conjunto de etiquetas es establecida según la estructura ordenada del conjunto de etiquetas [4], considerando que cada etiqueta del par (s_i, s_{T-i}) es igualmente informativa. Por ejemplo, podríamos usar el siguiente conjunto de 9 etiquetas para representar la información lingüística:

$$S = \{N, EL, VL, L, M, H, VH, EH, T\}.$$

Para operar con información lingüística ordinal es necesario definir una serie de operadores:

1. Negación: $Neg(s_i) = s_j, j = T - i$.
2. Comparación:
 - Maximización: $MAX(s_i, s_j) = s_i$ if $s_i \geq s_j$.
 - Minimización: $MIN(s_i, s_j) = s_i$ if $s_i \leq s_j$.
3. Agregación: Generalmente para combinar información lingüística ordinal usamos operadores de agregación basados en computación simbólica, como por ejemplo el operador LOWA [11].

3 Estructuras de los MOEAs con PG

3.1 Componentes

Los tres MOEAs adaptados al uso de árboles sintácticos de expresión estudiados en esta contribución comparten los siguientes componentes:

Esquema de codificación: las consultas son representadas mediante árboles sintácticos de expresión donde los nodos hoja representan los términos índice y sus ponderaciones lingüísticas y los nodos intermedios los conectivos lógicos AND y OR. Las ponderaciones lingüísticas son modeladas siguiendo el modelo lingüístico difuso ordinal.

Operador de Cruce: aleatoriamente se elige una rama de cada uno de los padres y se intercambian los subárboles de estas ramas entre ambos padres.

Operador de Mutación: la mutación es necesaria para variar las ponderaciones asociadas a los términos. Esta consiste en escoger de manera aleatoria el término a mutar y aumentar o disminuir el valor de ponderación asociado.

Generación de la población inicial: Todos los individuos de la primera generación son generados de manera aleatoria. La población es creada con todos los términos incluidos en el conjunto de documentos relevantes proporcionados por el

usuario, aquellos que aparezcan en más documentos tendrán mayor probabilidad de ser seleccionados.

Objetivos a optimizar: precisión y exhaustividad.

3.2 NSGA-II-PG

NSGA II [10] es un algoritmo muy completo ya que, no sólo incorpora una estrategia de preservación de la elite, sino que usa además un mecanismo explícito para preservar la diversidad. NSGA II trabaja con una población de descendientes Q_t que se crea usando la población padre P_t . Ambas poblaciones (Q_t y P_t) se combinan para formar una sola llamada R_t , de tamaño $2 \cdot M$, de la que se busca extraer el frente del Pareto. Es entonces cuando se realiza una ordenación sobre los individuos no-dominados para clasificar la población R_t . Aunque esto supone un mayor esfuerzo si lo comparamos únicamente con la ordenación del conjunto Q_t , permite una comprobación global de las soluciones no-dominadas que pertenecen tanto a la población de descendientes como a la de los padres.

Una vez que la ordenación de los individuos no-dominados termina, la nueva generación se forma con soluciones de los diferentes frentes de no-dominados, tomando de uno de los frentes cada vez. Se comienza con el mejor frente de individuos no-dominados y se continúa con las soluciones del segundo frente, después con el tercero, etc. Como el tamaño de R_t es $2 \cdot M$, no todos los frentes pueden pertenecer a la nueva población, ya que ésta tiene sólo tamaño M . Todos los frentes de soluciones que no pasan a la nueva población se eliminan directamente.

Durante las últimas etapas de la ejecución, es probable que la mayoría de las soluciones de la población se encuentren en el mejor frente de no-dominadas. También es probable que el número de soluciones en el primer frente de no-dominadas de la población combinada R_t (de tamaño $2 \cdot M$) sea mayor que M .

Es en ese momento cuando el algoritmo anterior asegura la selección de un conjunto diverso de soluciones de este frente mediante el método de nichos. Cuando la población entera converge a la frontera Pareto-óptima, el algoritmo continúa, de forma que se asegure la mejor distribución entre las soluciones.

3.3 SPEA2-PG

SPEA2 [27] introduce elitismo por el mantenimiento explícito de una población externa \bar{P} . Esta población almacena un número fijo de soluciones no-dominadas encontradas desde el comienzo de la simulación.

En cada generación, las nuevas soluciones no-

dominadas encontradas se comparan con la población externa existente y se preservan las soluciones no-dominadas resultantes. Además, SPEA2 usa estas soluciones elite para participar en las operaciones genéticas con la población actual con la esperanza de influenciar a la población para conducirla hacia buenas regiones en el espacio de búsqueda.

El algoritmo comienza con una población P_0 de tamaño M que se crea de forma aleatoria y una población externa \bar{P}_0 que en principio se encuentra vacía y tiene un máximo de capacidad \bar{M} . En cada generación t , las mejores soluciones no-dominadas (pertenecientes al primer frente no-dominado) de las poblaciones P_t y \bar{P}_t se copian en la población externa \bar{P}_{t+1} . Si el tamaño de \bar{P}_{t+1} excede \bar{M} entonces se reduce \bar{P}_{t+1} mediante un operador de truncamiento, en caso contrario \bar{P}_{t+1} se completa hasta \bar{M} con soluciones dominadas de P_t y \bar{P}_t . Este operador de truncamiento mantiene la diversidad.

A partir de \bar{P}_{t+1} obtenemos un pool de individuos aplicando un selección por torneo binario con remplazamiento. Estos individuos son cruzados y mutados para obtener la nueva generación P_{t+1} .

3.4 MOGLS-PG

MOGLS es un enfoque híbrido, que combina conceptos de MOEAs y Local Search para mejorar la población actual. En la literatura especializada se han propuesto varias alternativas [12, 13, 14, 15]. De todas ellas, hemos optado por estudiar el rendimiento de la primera propuesta de Ishibuchi [12]. Este enfoque centra gran parte de su esfuerzo en conseguir una frontera Pareto lo más extendida posible, para ello, asocia a cada individuo de la población un vector de pesos que marcan la dirección, en el espacio objetivo, con la que dicho individuo fue generado. Estas direcciones son tenidas en cuenta a la hora de generar nuevos individuos.

Este algoritmo mantiene el elitismo usando dos conjuntos de soluciones: la población actual y una población provisional de soluciones no-dominadas.

El algoritmo empieza generando una población inicial de N_{pop} individuos, los evalúa y actualiza la población provisional con los individuos no-dominados obtenidos. A continuación obtiene un conjunto de $N_{pop} - N_{elite}$ padres. Estos padres serán seleccionados teniendo en cuenta su vector dirección. Después del correspondiente cruce y mutación, la población será completada con N_{elite} individuos no-dominados de la población provisional. A continuación se aplica una búsqueda local a todos los N_{pop} individuos de la población actual. En esta etapa el vector dirección de cada individuo guiará el proceso de búsqueda local.

Finalmente, la generación siguiente es obtenida con los N_{pop} individuos mejorados de la población actual.

3.5 Evaluación de MOEAs

Cuando trabajamos con optimización multiobjetivo, la definición de calidad es substancialmente más compleja que con problemas de optimización mono-objetivo, ya que los procesos de optimización en sí implican varios objetivos. En la literatura especializada se han propuesto diferentes medidas cuantitativas [7, 9, 16, 26]. Una de las más utilizadas es la medida C cuya expresión:

$$C(A, B) = \frac{|\{a \in A : \exists b \in B : b \succ a\}|}{|A|}$$

mide el porcentaje de individuos del pareto A que son dominados por individuos del pareto B . Un valor de 1 indicaría que todos los individuos del pareto A son dominados por individuos del pareto B ; por su parte, un valor de 0 nos dice que ninguno de los individuos de A es dominado por individuos de B .

4 Estudio experimental

Como se ha comentado, el estudio experimental se ha desarrollado utilizando la colección Cranfield, compuesta por 1400 documentos sobre Aeronáutica. Los 1400 documentos de texto han sido automáticamente indizados de la forma usual, eliminando las palabras vacías, obteniendo un total de 3857 términos índice diferentes e indizando según el esquema $tf \cdot idf$ [20] para generar los términos relevantes en las representaciones de los documentos. De entre las 225 consultas asociadas a Cranfield, se han seleccionado aquellas que presentan 20 o más documentos relevantes. Las siete consultas resultantes (números 1, 2, 23, 73, 157, 220 y 225) tienen asociados 29, 25, 33, 21, 40, 20, 25 documentos relevantes, respectivamente.

Los MOEAs estudiados en esta contribución han sido ejecutados 30 veces para cada una de las consultas utilizadas, dando un total de 630 ejecuciones. De cada una de ellas se obtiene un pareto. Los tres paretos obtenidos por cada ejecución y consulta son comparados con la medida de rendimiento C (en la Tabla 1 se muestran los resultados promedio de C para cada pareja de EAs y consulta) y analizados de manera estadística mediante un test de Wilcoxon [22]. En la Tabla 2 mostramos los resultados de la comparación, donde el signo + indica que estadísticamente el algoritmo de la fila es mejor que el de la columna, el signo - indica lo contrario. Los parámetros generales utilizados son: 50.000 evaluaciones, 800 individuos por población, probabilidad de cruce = 0.8, probabilidad

Tabla 1: Resultado promedio de la medida $C(A, B)$ para cada una de las consultas estudiadas. X=NSGA-II-PG, Y=MOGLS-PG y Z=SPEA2-PG.

Consulta	C(X,Y) / C(Y,X)	C(X,Z) / C(Z,X)	C(Y,Z) / C(Z,Y)
1	0.831 / 0	1 / 0	0.008 / 0.960
2	0.754 / 0	1 / 0	0.025 / 0.920
23	0.728 / 0.005	0.985 / 0	0.013 / 0.920
73	0.705 / 0.028	0.951 / 0	0.031 / 0.917
157	0.727 / 0.011	0.975 / 0	0.002 / 0.953
220	0.768 / 0.011	0.970 / 0	0.011 / 0.932
225	0.782 / 0	0.974 / 0	0.002 / 0.981

de mutación = 0.2 y 20 nodos como máximo en el árbol sintáctico de expresión. Los parámetros específicos para MOGLS son: tamaño elite = 200, probabilidad de aplicar búsqueda local = 0.3 y máximo de 10 vecinos evaluados; y para SPEA2-PG: tamaño de la población élite = 200.

Tabla 2: Resultado del test de Wilcoxon, considerando $N = 7 : 2$

C(A, B)	MOGLS-PG	SPEA2-PG	NSGA-II-PG
NSGA-II-PG	+	+	-
SPEA2-PG	-	-	-
MOGLS-PG	-	+	-

5 Conclusiones

En esta contribución hemos desarrollado un estudio comparativo del desempeño, aprendiendo consultas lingüísticas difusas, de tres de los MOEAs más utilizados recientemente en la literatura. Para ello, hemos desarrollado e implementado versiones PG de los citados MOEAs (NSGA-II-PG, SPEA2-PG, MOGLS-PG) y validado con la base de datos de test Cranfield. Los resultados arrojados por este estudio indican que NSGA-II-PG es el enfoque que mejor se comporta en esta faceta.

Referencias

- [1] R. Baeza-Yates and B. Ribeiro-Neto, *Modern information retrieval*, Addison-Wesley, 1999.
- [2] G. Bordogna, P. Carrara, and G. Pasi, *Fuzzy sets and possibility theory in database management systems*, ch. Fuzzy Approaches to Extend Boolean Information Retrieval, pp. 231-274, Bosc, P. and Kacprzyk, J., Eds. (Springer Verlag), 1995.
- [3] G. Bordogna and G. Pasi. *A fuzzy linguistic approach generalizing boolean information retrieval: A model and its evaluation*. Journal of the American Society for Information Science 44 (1993), no. 2, 70-82.

- [4] ———, *The ordered weighted averaging operators: Theory and applications*, ch. Application of the OWA operators to soften information retrieval systems, pp. 275–294, Yager, R.R. and Kacprzyk, J., Eds. (Kluwer Academic Publishers), 1997.
- [5] D. Buell and D.H. Kraft, *A model for a weighted retrieval system*, Journal of the American Society for Information Science **32** (1981), 211–216.
- [6] H. Chen, G. Shankaranarayanan, L. She, and A. Iyer, *A machine learning approach to inductive query by example: An experiment using relevance feedback, ID3, genetic algorithms, and simulated annealing*, Journal of the American Society for Information Science **49** (1998), no. 8, 693–705.
- [7] C.A. Coello, D.A. Van Veldhuizen, and G.B. Lamant, *Evolutionary algorithms for solving multi-objective problems*, Kluwer Academic Publishers, 2002.
- [8] O. Cordón, E. Herrera-Viedma, López-Pujalte C., and C. Luque, M. Zarco, *A review on the application of evolutionary computation to information retrieval*, International Journal of Approximate Reasoning **34** (2003), 241–264.
- [9] K. Deb, *Multi-objective optimization using evolutionary algorithms*, Wiley, 2001.
- [10] K. Deb, A. Pratap, S. Agrawal, and Meyarivan T., *A fast and elitist multiobjective genetic algorithm: NSGA-II*, IEEE Transactions on Evolutionary Computation **6** (2002), 182–197.
- [11] F. Herrera, E. Herrera-Viedma, and J.L. Verdegay, *Direct approach processes in group decision making using linguistic owa operators*, Fuzzy Sets and Systems **79** (1996), 175–190.
- [12] H. Ishibuchi and T. Murata, *A multi-objective genetic local search algorithm and its application to flowshop scheduling*, IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews **28** (1998), no. 3.
- [13] H. Ishibuchi, T. Yoshida, and T. Murata, *Balance between genetic search and local search in memetic algorithms for multiobjective permutation flowshop scheduling*, IEEE Transactions on Evolutionary Computation **7** (2003), no. 2, 204–223.
- [14] A. Jaskiewicz, *Do multiple-objective metaheuristics deliver on their promises? a computational experiment on the set-covering problem*, IEEE Transactions on Evolutionary Computation **7** (2003), no. 2, 133–143.
- [15] ———, *Genetic local search for multi-objective combinational optimization*, European Journal of Operational Research **137** (2003), 50–71.
- [16] J. Knowles, L. Thiele, and E. Zitzler, *A tutorial on the performance assessment of stochastic multiobjective optimizers*, 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland, feb 2006, revised version.
- [17] J. Koza, *Genetic programming. on the programming of computers by means of natural selection*, The MIT Press, 1992.
- [18] D.H. Kraft, F.E. Petry, B.P. Buckles, and T. Sadasivan, *Genetic algorithms and fuzzy logic systems*, ch. Genetic algorithms for query optimization in information retrieval: relevance feedback, pp. 155–173, Sanchez, E. and Shibata, T. and Zadeh, L.A., Eds., (World Scientific), 1997.
- [19] S. Miyamoto, *Fuzzy sets in information retrieval and cluster analysis*, Kluwer Academic Publishers, 1990.
- [20] G. Salton, *Automatic text processing: The transformation, analysis and retrieval of information by computer*, Addison-Wesley, 1989.
- [21] G. Salton and M.J. McGill, *An introduction to modern information retrieval*, McGraw-Hill, 1983.
- [22] D.J. Sheskin, *Handbook of parametric and non-parametric statistical procedures*, New York, NY: CRC Press, 1997.
- [23] M. P. Smith and M. Smith, *The use of genetic programming to build boolean queries for text retrieval through relevance feedback*, Journal of Information Science **23** (1997), no. 6, 423–431.
- [24] V. Tahani, *A fuzzy model of document retrieval systems*, Information Processing & Management **12** (1976), 177–187.
- [25] C.J. van Rijsbergen, *Information retrieval*, Butterworth, 1979.
- [26] E. Zitzler, K. Deb, and L. Thiele, *Comparison of multiobjective evolutionary algorithms: Empirical results*, Evolutionary Computation **8** (2000), no. 2, 173–271.
- [27] E. Zitzler, M. Laumanns, and L. Thiele, *SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization*, Evolutionary Methods for Design, Optimization and Control with Application to Industrial Problems (EUROGEN 2001) (K. Giannakoglou et al. (Eds.), ed.), 2002, pp. 95–100.

Aproximaciones simples a una superficie mediante multiresolución y TSK	197
Adolfo R. de Soto; Conrado A. Capdevila	
Hacia un cálculo fuzzy numérico/simbólico	205
Adolfo R. de Soto	
Uso de gráficos de cajas borrosos para controlar el final de una aplicación del método delphi	211
Sarabia, A., Maté, C., Caro, R. y J. Tapia	
An Evolutionary Fuzzy System for Scheduling in Wood-based Furniture Manufacturing ..	217
Juan C. Vidal Aguiar, M. Mucientes, M. Lama and A. Bugarín	
Verificación de Taxímetros usando mediciones GPS como conjuntos borrosos	223
Jose R. Villar, Luciano Sanchez, Adolfo Otero, Jose Otero	

Sesiones Especiales

Soft Computing para la recuperación de información y minería en Web

Organizadores: Enrique Herrera-Viedma, María J. Martín-Bautista, José Ángel Olivas, Alejandro Sobrino
(Red Temática Nacional sobre Sistemas de Acceso a la Información en la Web basados en Soft Computing).

Uso de modelos de restricción del contexto para el desarrollo de aplicaciones móviles inteligentes	229
Fernando Bobillo, Miguel Delgado, Juan Gómez-Romero.	
Minería Web: Aplicaciones con Lógica Difusa	235
V.H. Escobar Jeria, M.J. Martín-Bautista, D. Sánchez, M.A. Vila.	
Un enfoque deductivo para la minería de textos	241
C. Justicia de la Torre, M.J. Martín-Bautista, D. Sánchez, M.A. Vila	
Aprendiendo consultas lingüísticas difusas con técnicas multiobjetivo. Estudio preliminar	247
A.G. López-Herrera, F. Herrera, E. Herrera-Viedma, A. Zafra, M. Luque.	
Caracterización automática de perfiles de usuarios basados en consultas lingüísticas multigranulares usando un algoritmo genético multiobjetivo	253
M. Luque, E. Herrera-Viedma, O. Cordon, A. Zafra, A.G. Lopez.	
Mejorando la búsqueda web mediante la adaptación de consultas en GUMSe	261
Javier de la Mata, José A. Olivas, Jesús Serrano.	
Técnicas Difusas y No-Difusas para el Clustering del Contenido de Páginas en un Sitio Web	267
David Nettleton, Barbara Poblete	
Un sistema de recomendaciones lingüístico difuso para filtrar recursos en bibliotecas digitales universitarias	273
C. Porcel, A.G. Lopez, S. Alonso, M. Luque, E. Herrera-Viedma.	

Estylf 2006

**XIII Congreso Español sobre
Tecnologías y Lógica Fuzzy**

Organización
Grupo de Investigación ORETO
Escuela Superior de Informática
Universidad de Castilla-La Mancha



ESTYLF 2006
Ciudad Real
20/22 Septiembre



Este libro de actas recoge todos los trabajos que han sido presentados como contribuciones al *XIII Congreso Español de Tecnologías y Lógica Fuzzy (ESTYLF)* de la *European Society for Fuzzy Logic and Technology (EUSFLAT)*.

El XIII Congreso ESTYLF ha sido organizado por el grupo de investigación Oreto, del departamento de Tecnologías y Sistemas de Información de la Universidad de Castilla-La Mancha, en las instalaciones de la Escuela Superior de Informática en Ciudad Real.

Desde estas líneas queremos agradecer a todos los que han contribuido a la elaboración de estas actas, tanto en su aspecto visual como en sus contenidos. De igual forma agradecemos la colaboración de las instituciones que han participado y con su apoyo han permitido la celebración de este congreso científico bianual.

Esperando que este encuentro en Ciudad Real sirva para mostrar el potencial que esta joven Universidad de Castilla-La Mancha es capaz de generar, así como ser unos dignos representantes del carácter acogedor de esta tierra.

Entidades Colaboradoras



Universidad de Castilla-La Mancha
Vicerrectorado del Campus de
Ciudad Real y Extensión Universitaria
Vicerrectorado de Investigación



Castilla-La Mancha

Consejería de
Educación y Ciencia



Escuela Superior de
Informática

Departamento de
Tecnologías y Sistemas
de Información

tsi

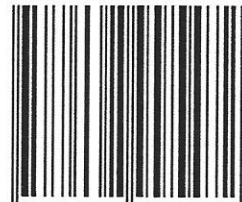


Diputación Provincial
de Ciudad Real

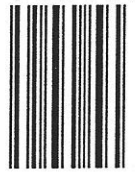
Ayuntamiento
de Ciudad Real



ISBN 84-689-9547-9



90000>



9 788468 995472

Editado por:

Grupo de Investigación Oreto.

Impreso en España por:

Artes Gráficas Lince

ISBN: 84-689-9547-9

N. REG: 06/50473