



---

# Índice general

---

<b>Resumen</b>	<b>XIII</b>
<b>Resum</b>	<b>XV</b>
<b>Abstract</b>	<b>XVII</b>
<b>Agradecimientos</b>	<b>XIX</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Objetivos . . . . .	4
1.3. Estructura del trabajo . . . . .	4
<b>I Estado del Arte</b>	<b>7</b>
<b>2. Técnicas Soft Computing</b>	<b>9</b>
2.1. Algoritmos Genéticos . . . . .	10
2.1.1. Codificación . . . . .	13
2.1.2. Operadores de Selección . . . . .	14
2.1.3. Operadores de Mutación . . . . .	18
2.1.4. Operadores de Cruce . . . . .	20
2.2. Redes Neuronales Artificiales . . . . .	24
2.2.1. Perceptrón Multicapa . . . . .	26
2.2.2. Redes Neuronales de Funciones Base Radiales . . . . .	27
2.2.3. Mapas auto-organizativos . . . . .	29
2.3. Sistemas de Lógica Difusa . . . . .	29
2.4. Máquinas de Soporte Vectorial . . . . .	30
2.5. Conclusiones . . . . .	33

<b>3. Catálisis Combinatoria</b>	<b>37</b>
3.1. Introducción . . . . .	37
3.2. IA aplicada a la Catálisis Combinatoria . . . . .	40
3.3. Conclusiones . . . . .	42
<b>4. Sistemas de Recomendación</b>	<b>45</b>
4.1. Modelo de recomendación . . . . .	47
4.2. Preferencias de los usuarios . . . . .	50
4.3. Importancia de la ontología empleada . . . . .	52
4.4. Escasez de Información . . . . .	54
4.5. Escalabilidad y rendimiento . . . . .	55
4.6. Algoritmos de Recomendación . . . . .	55
4.6.1. Métodos basados en el contenido . . . . .	56
4.6.2. Filtrado por colaboración basada en los ítems . . . . .	59
4.6.3. Filtrado por colaboración basada en los usuarios . . . . .	61
4.6.4. Métodos Híbridos . . . . .	65
4.7. Conclusiones . . . . .	69
<b>II Propuesta</b>	<b>73</b>
<b>5. Arquitectura propuesta</b>	<b>75</b>
5.1. Configuración . . . . .	76
5.1.1. Codificación del problema . . . . .	77
5.1.2. Configuración de los parámetros Soft Computing . . . . .	80
5.1.3. Obtención de la Generación Inicial . . . . .	82
5.1.4. Obtención del modelo de RNA . . . . .	83
5.2. Re-entrenamiento de la RNA . . . . .	84
5.3. Modelado de la función de fitness . . . . .	85
5.4. Aplicación de los operadores del AG . . . . .	85
5.5. Evaluación de simulación . . . . .	90
5.6. Evaluación de control . . . . .	90
5.7. Conclusiones . . . . .	91
<b>6. Evaluación y Resultados</b>	<b>93</b>
6.1. Aplicación a la Catálisis Combinatoria . . . . .	94
6.1.1. Redes Neuronales modelando resultados catalíticos . . . . .	94
6.1.2. Obtención de bibliotecas de modelos . . . . .	103
6.1.3. Convergencia del Algoritmo Genético al emplear modelos aproximados de las funciones de aptitud . . . . .	117
6.1.4. Evaluación de las etapas propuestas en la arquitectura . . . . .	129
6.1.5. Aplicación a reacciones de interés industrial . . . . .	156
6.2. Aplicación a los Sistemas de Recomendación . . . . .	162
6.2.1. Configuración de los parámetros de la arquitectura determinando las preferencias de un usuario. . . . .	164

6.2.2.	Determinación de las preferencias de grupos de usuarios similares identificados mediante correlación de Pearson . . . . .	171
6.2.3.	Determinación de las preferencias de grupos de usuarios similares empleando modelos aproximados de la función de aptitud . . . . .	178
<b>7.</b>	<b>Conclusiones</b>	<b>187</b>
7.1.	Aportaciones . . . . .	187
7.2.	Lineas Futuras de Investigación . . . . .	191
7.3.	Publicaciones . . . . .	192
<b>A.</b>	<b>Paquete SoftCombi</b>	<b>195</b>
A.1.	Herramienta DoE para la Catálisis Combinatoria . . . . .	195
A.1.1.	Principales Funcionalidades . . . . .	197
A.2.	Herramienta para la obtención de modelos . . . . .	198
	<b>Bibliografía</b>	<b>201</b>



# Índice de figuras

2.1.	Esquema de actuación de un algoritmo genético simple . . . . .	11
2.2.	Selección Proporcional. Métodos de ruleta y selección estocástica universal (SEU) . . . . .	16
2.3.	Selección por Ranking. Métodos de Ranking y Rank-Space, tomando para el cálculo de la probabilidad de selección: $min = 0.5$ y $max = 1.5$ . . . . .	17
2.4.	Ejemplo de aplicación de operadores de cruce y mutación básicos sobre cromosomas binarios . . . . .	18
2.5.	Perceptron Multicapa, con una capa de entrada, una oculta y una de salida	27
2.6.	Red Neuronal de funciones base radial . . . . .	28
2.7.	Aplicación de una función de mapeo $\varphi(\cdot)$ para transformar las muestras a fin de que sean separables mediante un hiperplano . . . . .	31
2.8.	Hiperplano separador encontrado durante el entrenamiento de una SVM	33
5.1.	Etapas de la arquitectura de búsqueda <i>Soft Computing</i> propuesta . . . . .	77
5.2.	Esquema de codificación de un cromosoma . . . . .	79
5.3.	Ejemplo de codificación para una formulación general de un catalizador basado en oro . . . . .	80
5.4.	Pasos a realizar durante la etapa de re-entrenamiento de la RNA . . . . .	84
5.5.	Ejemplos de la operación de mutación por cambio de selección . . . . .	87
5.6.	Ejemplos de las operaciones de mutación y cruce . . . . .	89
6.1.	RNA entrenada con muestras de generaciones previas empleada para predecir los resultados catalíticos de la generación actual . . . . .	96
6.2.	Resultados del proceso de entrenamiento . . . . .	98
6.3.	Resultados del proceso de test . . . . .	98
6.4.	Resultados experimentales y aproximados para el rendimiento del etano y la conversión del $O_2$ . . . . .	100
6.5.	Evolución de las predicciones ofrecidas por el modelo habiendo sido entrenado con muestras procedentes de las generaciones anteriores. . . . .	102
6.6.	Esquema de isomerización seguido por las parafinas $n-C_8$ , $n-C_7$ , $n-C_6$ y $n-C_5$ . . . . .	103

6.7. Esquema seguido en el proceso de obtención de bibliotecas de modelos mediante re-entrenamiento. . . . .	106
6.8. Media del error absoluto cometido por las mejores RNA al modelar los resultados catalíticos de conversión para la reacción n-octano. . . . .	108
6.9. Media del error cuadrático medio cometido por las mejores RNA al modelar los resultados catalíticos de conversión para la reacción n-octano. . . . .	108
6.10. Topología de red seleccionada como mejor modelo para las reacciones n-parafinas. . . . .	109
6.11. Predicciones de conversión obtenidas para la reacción n-octano mediante la RNA 4-8-6-3. . . . .	109
6.12. Predicciones del rendimiento a mono-ramificados obtenidas para la reacción n-octano mediante la RNA 4-8-6-3. . . . .	110
6.13. Predicciones del rendimiento a di-ramificados obtenidas para la reacción n-octano mediante la RNA 4-8-6-3. . . . .	110
6.14. Error cuadrático medio cometido en la predicción de la conversión para la reacción $n-C_8$ con la RNA 4-8-6-3 empleando la función de activación logística . . . . .	111
6.15. Error cuadrático medio cometido en la predicción de la conversión para la reacción $n-C_8$ con la RNA 4-8-6-3 empleando la función de activación tangencial. . . . .	112
6.16. Representación de los errores de predicción medios normalizados para la conversión y rendimientos para mono y di-ramificados . . . . .	113
6.17. Influencia del número de muestras de los ConjB en los resultados ofrecidos por los modelos para predecir el rendimiento a di-ramificados. . . . .	114
6.18. Predicciones del rendimiento a di-ramificados para la reacción n-heptano empleando una red $RNA_{C_7(85)}$ . . . . .	114
6.19. Predicciones de la conversión alcanzada para 300 muestras de test empleando las redes $RNA_{C_6C_8(850,10)}$ y $RNA_{C_6C_7(850,10)}$ . . . . .	116
6.20. Predicciones de la conversión, rendimiento a mono y di-ramificados empleando la $RNA_{C_8C_7(850,10)}$ con y sin emulación de error experimental. . . . .	118
6.21. Combinación de un AG y una RNA para la optimización de las condiciones de reacción de los procesos de isomerización de n-parafinas. . . . .	121
6.22. Evolución de la calidad media durante el estudio realizado para determinar la mejor PM y NumGen más adecuados. . . . .	125
6.23. Resultados obtenidos en el estudio realizado para determinar la mejor combinación de PC) y Pr. . . . .	126
6.24. Resultados obtenidos en el estudio realizado para determinar la mejor combinación de los parámetros $\alpha$ y SC . . . . .	127
6.25. Representación de la función hipotética en los planos que contienen los máximos absolutos . . . . .	131
6.26. Codificación del problema de optimización de un catalizador para la reacción ODHP . . . . .	134
6.27. Aproximaciones de la conversión y selectividad realizadas por el modelo 5_10_2 entrenado y probado con conjuntos de muestras de diferente tamaño. . . . .	137

6.28. Aproximaciones del rendimiento obtenidas mediante el modelo 5_10_2 para muestras con error experimental emulado . . . . .	138
6.29. Evolución de la aptitud media de las generaciones para cada combinación de parámetros del operador de mutación estudiados. . . . .	139
6.30. Evolución de la aptitud media de las generaciones para cada combinación de parámetros del operador de cruce estudiados . . . . .	139
6.31. Evolución de la aptitud media de las generaciones al variar el tamaño de la población estudiada. . . . .	140
6.32. Evolución de la aptitud media de las generaciones al optimizar la composición para un catalizador para la reacción de deshidrogenación oxidativa del propano. . . . .	141
6.33. Aproximaciones realizadas por la red en la etapa de re-entrenamiento vs. valores experimentales. . . . .	143
6.34. Influencia del tamaño de la población. Valores medios y máximos de la aptitud obtenida. . . . .	144
6.35. Influencia del método de actualización del modelo basado en una RNA en la precisión final del mismo . . . . .	146
6.36. Efecto de emplear evaluación simulada. Aptitudes máximas obtenidas con distintos tamaños de poblaciones virtuales y ratios de reducción, consiguiendo generaciones de control de 35, 45 y 55 muestras. . . . .	149
6.37. Máximos alcanzados y número de muestras de control evaluadas por ciclo de optimización realizado, utilizando poblaciones virtuales de entre 55 y 75 muestras. . . . .	151
6.38. Correlación entre los máximos obtenidos y el número de muestras de control evaluadas (coste experimental). . . . .	152
6.39. Comparativa de los resultados obtenidos al utilizar diferentes estrategias de búsqueda en las que se evalúa una población de control de 35 individuos.157	
6.40. Comparativa de los resultados conseguidos al utilizar diferentes estrategias de búsqueda en las que se obtiene una población de control de 45 individuos. . . . .	157
6.41. Codificación del problema de optimización del catalizador Ti-silicato . .	158
6.42. Evolución de la aptitud de las muestras del catalizador Ti-silicato para la epoxidación de oleofinas. . . . .	160
6.43. Resultados de epoxidación para los mejores catalizadores encontrados en cada generación a dos temperaturas de test diferentes. . . . .	161
6.44. Codificación del problema MoviLens. Define el conjunto de pesos sobre los atributos de las películas que describen las preferencias de los usuarios.165	
6.45. Evolución de la aptitud media alcanzada en cada generación propuesta por el AG en la fase de entrenamiento . . . . .	170
6.46. Conjunto de pesos para los diferentes atributos de las películas (cromosoma) que define las preferencias del usuario 276 del sistema <i>MoviLens</i> . 170	
A.1. Herramienta para el diseño inteligente de experimentos en el ámbito de la Catálisis Combinatoria basada en técnicas Soft Computing . . . . .	196

A.2. Herramienta para la obtención de modelos basados en perceptrones multicapa . . . . .	199
-------------------------------------------------------------------------------------------	-----



# Índice de tablas

2.1. Ventajas e Inconvenientes de técnicas empleadas en sistemas <i>Soft Computing</i> . . . . .	34
4.1. Ventajas e inconvenientes de los principales algoritmos de recomendación	70
6.1. Media de los errores absolutos cometidos por la RNA al predecir los resultados catalíticos de las diferentes generaciones ofrecidas por el AG .	101
6.2. Distribución de los resultados catalíticos para las reacciones n-octano, n-heptano, n-hexano, n-pentano empleados para obtener los modelos basados en perceptrones multicapa. . . . .	105
6.3. Topologías, funciones de activación y conjuntos de muestras de entrenamiento empleadas en el estudio realizado para la obtención de una topología de perceptrón adecuada para reacciones de n-parafinas. . . . .	106
6.4. Parámetros del AG y valores que se estudiarán para determinar la mejor combinación . . . . .	119
6.5. Evolución de la calidad media obtenida por la combinación del AG y la RNA cuando la generación inicial (generación 0) tiene una baja calidad contra la calidad obtenida cuando la generación de partida es de muy buena calidad. . . . .	128
6.6. Resultados obtenidos en la generación 15 para cada combinación de RNA y AG probadas: calidades medias, máximas y errores absolutos cometidos en la predicción de los resultados catalíticos. . . . .	129
6.7. Comparativa de las calidades alcanzadas por los cinco mejores y peores individuos de la última generación obtenidos para la reacción de n-octano al emplear redes con distinto grado de conocimiento. . . . .	129
6.8. Topologías de RNA, algoritmos de entrenamiento y conjuntos de muestras a estudiar . . . . .	133
6.9. Valores a estudiar para los diferentes parámetros de la arquitectura de búsqueda requeridos en la etapa IV . . . . .	135
6.10. Topologías de RNA con mejor comportamiento para los conjuntos de muestras <i>Set1</i> y <i>Set2</i> . . . . .	136

6.11. MAE y MSE obtenidos por las RNA con mejores resultados al modelar la función hipotética 6.3. . . . .	136
6.12. MAE y MSE obtenidos por las RNA con mejores resultados al modelar la función hipotética 6.7. . . . .	142
6.13. Valores estudiados para los parámetros relacionados con la etapa de evaluación simulada de la arquitectura. . . . .	148
6.14. Parámetros de la arquitectura de búsqueda <i>Soft Computing</i> utilizados para evaluar la utilidad de la obtención de <i>generaciones internas</i> . . . .	154
6.15. Comparativa de las muestras de control empleadas con respecto a las muestras exploradas, así como la aptitud máxima alcanzada tras finalizar diez ciclos de optimización en las diferentes pruebas realizadas. . . . .	156
6.16. Valores utilizados en el estudio realizado para encontrar la mejor combinación para los parámetros relacionados con el operador de cruce y el tamaño de la población explorada. . . . .	167
6.17. Valores estudiados en la búsqueda de la mejor combinación para los parámetros relacionados con el operador de mutación. . . . .	167
6.18. Mejores resultados obtenidos en la fase de test durante la optimización de los parámetros de cruce y población. . . . .	168
6.19. Mejores resultados obtenidos en la fase de test durante la optimización de los parámetros de mutación. . . . .	169
6.20. Usuarios que conforman los grupos de usuarios similares teniendo en cuenta las 3 ó 5 primeras valoraciones realizadas por los usuarios de referencia, ordenados de mayor a menor similitud. . . . .	174
6.21. Resultados obtenidos para cada unos de los grupos <i>userId-xM-xxk</i> al estimar las valoraciones presentes en los corpus de test. . . . .	175
6.22. Resultados obtenidos al emplear las preferencias aprendidas para cada unos de los grupos <i>userId-xM-xxk</i> al estimar las valoraciones de los usuarios de referencia. . . . .	176
6.23. Resultados de aplicar las preferencias aprendidas para los grupos de usuarios 276-3M-xxk para estimar valoraciones de usuarios similares al 276. . . . .	177
6.24. Resultados de aplicar las preferencias aprendidas para los grupos de usuarios 276-5M-xxk para estimar valoraciones de usuarios similares al 276. . . . .	177
6.25. Resultados de aplicar las preferencias aprendidas para los grupos de usuarios 710-xM-xxk para estimar valoraciones de usuarios similares al 710. . . . .	178
6.26. Valores estudiados para los diferentes aspectos considerados en la selección de una topología de perceptrón y función de entrenamiento adecuados	179
6.27. Mejores resultados obtenidos durante el estudio realizado para la obtención de un modelo capaz de ofrecer aproximaciones de valoraciones. . . .	182
6.28. Resultados obtenidos al emplear las preferencias aprendidas para el grupo de usuarios 276-5M-20k para estimar las valoraciones de diferentes corpus de test. . . . .	184

6.29. Comparativa de las preferencias aprendidas para el grupo 276-5M-20k al emplear o no valores aproximados de la función de aptitud. . . . . 184