Estudio de un Modelo de Ensemble formado por los Extremos del Frente de Pareto en Algoritmos Evolutivos Multi-objetivo Utilizando la Técnica TOPSIS

M. Cruz-Ramírez, C. Hervás-Martínez
Dept. de Informática y Análisis Numérico
Escuela Politécnica Superior
Universidad de Córdoba
14071 Córdoba
i42crram@uco.es, chervas@uco.es

M. T. Lamata
Dept. Ciencias de la Computación e Inteligencia Artificial
ETS Ingeniería Informática y Telecomunicación
Universidad de Granada
18071 Granada
mtl@decsai.ugr.es

Resumen

Una tarea compleja y de contexto dependiente, es determinar qué modelo perteneciente al frente de Pareto es el que debemos usar en una toma de decisión. Por ello, en el presente trabajo proponemos el empleo de la técnica de decisión TOPSIS (*Technique for Order Preference by Similarity to Ideal Solution*) para ayudarnos en esta elección. Concretamente, con la ayuda del método TOPSIS determinaremos qué *ensemble* formado por la combinación lineal de los mejores modelos asociados a los extremos del frente de Pareto es la que produce mejor rendimiento en clasificación.

1. Introducción

En el mundo real existen multitud de problemas de multiclasificación, por lo que existe un creciente interés por ellos en la comunidad de *machine learning* [8]. Generalmente, un algoritmo diseña un clasificador para aproximar una función de entrada-salida desconocida a partir de un conjunto finito de datos disponibles, es decir, los patrones de entrenamiento. Una vez que este clasificador ha sido diseñado, puede utilizarse para predecir la clase de pertenencia de nuevos patrones, es decir, los patrones del conjunto de generalización. Por lo tanto, el objetivo es diseñar un buen clasificador que garantice una alta precisión en la predicción de futuros datos, es decir, con buena capacidad de generalización. Muchas técnicas han sido propuestas para mejorar la capacidad de generalización del clasificador diseñado [6] (utilizando como medida de rendimiento, por ejemplo, la maximización del porcentaje de patrones correctamente clasificados), pero muy pocos métodos utilizan como medida la maximización de ese porcentaje para cada una de las clases. Este segundo objetivo es muy importante en algunas áreas de investigación (como la medicina, la teledetección, la economía, etc.). Por tanto, es necesario realizar una optimización simultánea de los dos objetivos (en general, en conflicto en problemas reales). El método de optimización utilizado en este trabajo es un método multi-objetivo, y la principal diferencia con la optimización mono-objetivo es que el primero no suele proporcionar una solución al problema, sino un conjunto de soluciones que son igualmente buenas. Partiendo de este conjunto de soluciones al problema, las cuales forman un frente de Pareto, proponemos una agregación de los modelos que se encuentren en los extremos de dicho frente. Esto es, el mejor modelo obtenido medido en precisión y el mejor modelo obtenido medido en sensibilidad, siendo ésta la menor de las sensibilidades de cada una de las clases que componen un determinado problema. Esta agregación será una combinación lineal de la salida de los dos modelos, que proporcionará la clase de pertenencia de los patrones.

El problema ahora es decidir cuál es la mejor combinación lineal de las salidas de estos modelos. Para resolverlo se aplicará, con distintos valores de λ (parámetro que tomará valores en el intervalo $[0, 1]$ con un incremento de 0.1), la combinación lineal $\lambda * S_{ES} + (1 - \lambda) * S_{EI}$ (donde S_{ES} y S_{EI} son las salidas de los modelos que se encuentran en el extremo superior y en el extremo inferior del frente de Pareto, respectivamente) y se seleccionará la mejor alternativa mediante la técnica TOPSIS.

La estructura del artículo es la siguiente: La sección 2 expone las métricas de precisión y de mínima sensibilidad; la sección 3 presenta el método de ayuda a la decisión TOPSIS; la sección 4 explica los experimentos que se han realizado; en la sección 5 se aplica el método TOPSIS a los extremos del frente de Pareto; finalmente la sección 6 describe las principales conclusiones así como los trabajos futuros.

2. Precisión y mínima sensibilidad

En esta sección presentamos las dos métricas utilizadas para evaluar un clasificador: el porcentaje de patrones correctamente clasificados o Precisión (C) y la Mínima Sensibilidad (MS). Para evaluar un clasificador tradicionalmente se ha empleado la métrica C. Sin embargo, al usar C no podemos captar los diferentes aspectos del comportamiento de dos clasificadores en problemas de multiclasificación. Es por ello que en este trabajo se consideran estas dos medidas: la Precisión, $C = \frac{1}{N}\sum_{j=1}^{Q} n_{jj}$ (donde Q es el número de clases, N es el número de patrones en los conjuntos de entrenamiento o generalización y n_{jj} es el número de patrones de la clase j-ésima que son correctamente clasificados por el clasificador), y la Mínima Sensibilidad de todas las clases (MS), esto es, el menor porcentaje de patrones correctamente clasificados de cada clase, S_i, con respecto al número total de patrones en dicha clase, $MS = \min\{S_i\}$ para $i = 1, ..., Q$. La medida bidimensional (MS, C) expresa dos características asociadas con un clasificador: el rendimiento global (C) y el desempeño de la clase peor clasificada (MS). La elección de MS como una medida complementaria de C está justificada al considerar que C es la media ponderada de las sensibilidades de cada una de las clases. Para una descripción más detallada de estas medidas ver [4].

A priori, podríamos pensar que los objetivos MS y C pueden estar correlados positivamente. Esto puede ser cierto para valores pequeños de MS y C, pero en general no ocurre con valores cercanos a 1. Tampoco ocurre en problemas no balanceados en el número de patrones en cada clase ni en problemas con múltiples clases. Este hecho justifica el uso de un Algoritmo Evolutivo Multi-Objetivo (MOEA) [9].

Gracias al MOEA obtenemos un conjunto de soluciones (modelos de red neuronal) al problema de clasificación en entrenamiento, que serán usadas a la hora de generalizar. Estas soluciones se sitúan en el frente de Pareto, siendo el extremo superior (ES) de este frente el modelo con mayor número de patrones correctamente clasificados y el extremo inferior (EI) el modelo con mejor valor en Mínima Sensibilidad. Con estos dos modelos formaremos un conjunto de clasificadores o *ensemble*, cuya salida será una combinación lineal de las salidas asociadas a cada uno de los modelos.

3. La técnica TOPSIS

La técnica TOPSIS (*Technique for Order Preference by Similarity to Ideal Solution*) es uno de los métodos más utilizados en problemas de decisión multi-criterio. TOPSIS fue propuesto por Hwang y Yoon en [7]. La lógica subyacente de la técnica TOPSIS es definir la solución ideal positiva y la solución ideal negativa. La solución ideal positiva es la solución que maximiza los criterios de beneficio y reduce al mínimo los criterios de costes; la solución ideal negativa es aquella que maximiza los criterios de costes y reduce al mínimo los criterios de beneficio. La alternativa óptima es la que esté más cerca de la solución ideal positiva y más alejada de la solución ideal negativa. Las distintas alternativas se ordenan según su valor de "proximidad con la solución ideal", siendo la mejor alternativa aquella con mayor valor de proximidad. Un ejemplo de aplicación de la

Tabla 1: Características de los conjuntos de datos

Dataset	Patrones	Patrones entren.	Patrones general.	Variables entrada	Clases	Patrones por clase
BreastC	286	215	71	15	2	(201,85)
Pima	768	576	192	8	2	(500,268)
AustralianC	690	517	173	51	2	(307,383)
H-Statlog	270	202	68	13	2	(150,120)
Ionosphere	351	263	88	34	2	(126,225)
Vote	435	326	109	16	2	(267,168)
Iris	150	111	39	4	3	(50,50,50)
Balance	625	469	156	4	3	(288,49,288)
Dermatology	366	275	91	34	6	(112,61,72,49,52,20)

técnica TOPSIS multi-criterio a un problema de decisión se puede ver en [5].

4. Experimentos

Para el diseño experimental se han considerado nueve conjuntos de datos tomados del repositorio de la UCI [2]. El diseño se ha realizado utilizando un proceso de validación cruzada *holdout* con 30 ejecuciones, donde aproximadamente el 75 % de los patrones fueron seleccionados aleatoriamente como conjunto de entrenamiento y el 25 % restante como conjunto de generalización.

En la Tabla 1 se muestran las principales características de cada uno de los conjuntos de datos empleados en la experimentación: el número total de patrones para cada conjunto de datos, el número de patrones de entrenamiento y de generalización, el número de variables de entrada, el número de clases (salidas) y el número total de patrones por clase.

El MOEA utilizado está descrito en [3]. Este algoritmo está basado en la Evolución Diferencial [10] y en los trabajos de H. Abbass [1]. En este trabajo no se empleará la búsqueda local. Esto se debe a que se quiere analizar cuál es la mejor combinación de los extremos del frente, sin que estos se vean afectados por la búsqueda local o por cualquier otro procedimiento (*clustering*, *resampling*...).

5. Aplicación del método TOPSIS

A continuación aplicaremos el método TOPSIS con el fin de determinar la mejor combinación lineal de las salidas de los modelos extremos del primer frente de Pareto según [3].

Lo primero que debemos hacer es fijar las posibles soluciones al problema (alternativas) y los criterios de decisión que se emplearán. Las alternativas se seleccionarán a partir de la expresión $\lambda * S_{ES} + (1-\lambda) * S_{EI}$, donde λ es un parámetro que toma valores en el intervalo $[0,1]$, con un incremento de 0.1 y S_{ES} y S_{EI} son las salidas de los modelos que se encuentran en el extremo superior y en el extremo inferior del primer frente de Pareto, respectivamente. Por tanto, tendremos 11 alternativas diferentes, $A = A_1, ..., A_{11}$, donde A_1 se corresponde con el extremo superior del frente de Pareto en entrenamiento ($\lambda = 1$) y A_{11} con el extremo inferior del frente ($\lambda = 0$).

Como criterios se empleará la media de C, la desviación estándar de C, la media de MS y la desviación estándar de MS obtenidas tras ejecutar 30 veces el algoritmo. Por tanto, al trabajar con 9 conjuntos de datos y los mismos 4 criterios en cada una de ellas, consideraremos 36 criterios de decisión, $C = C_1, ..., C_{36}$, donde $C_1 = \widehat{\mu}_C$, $C_2 = \widehat{\sigma}_C$, $C_3 = \widehat{\mu}_{MS}$ y $C_4 = \widehat{\sigma}_{MS}$ son los estadísticos del conjunto de BreastC, los siguientes cuatro criterios se corresponden con los valores de los estadísticos para el conjunto de Pima y así sucesivamente hasta terminar con los valores de los estadísticos para Dermatology.

De estos criterios, lo deseable es maximizar los valores de $\widehat{\mu}_C$ y $\widehat{\mu}_{MS}$ y minimizar $\widehat{\sigma}_C$ y $\widehat{\sigma}_{MS}$ para todos los conjuntos de datos.

Una vez determinadas las alternativas y los criterios de decisión, se construye la matriz de decisión X.

Tabla 2: Matriz de decisión X obtenida sobre el conjunto de generalización

		A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}
BreastC	C_1	0.6667	0.6662	0.6657	0.6615	0.6582	0.6596	0.6615	0.6620	0.6502	0.6474	0.6366
	C_2	0.0317	0.0310	0.0327	0.0299	0.0282	0.0294	0.0258	0.0274	0.0335	0.0358	0.0359
	C_3	0.3810	0.3857	0.3905	0.3937	0.3984	0.4143	0.4302	0.4524	0.4959	0.5531	0.5632
	C_4	0.1263	0.1284	0.1278	0.1275	0.1326	0.1270	0.1278	0.1217	0.0918	0.0782	0.0750
Pima	C_5	0.7684	0.7681	0.7668	0.7658	0.7649	0.7634	0.7628	0.7615	0.7599	0.7569	0.7550
	C_6	0.0305	0.0307	0.0296	0.0282	0.0281	0.0295	0.0275	0.0289	0.0304	0.0315	0.0312
	C_7	0.6075	0.6204	0.6303	0.6423	0.6522	0.6647	0.6761	0.6876	0.6945	0.7047	0.7103
	C_8	0.0705	0.0669	0.0640	0.0598	0.0517	0.0463	0.0445	0.0475	0.0495	0.0501	0.0483
AustralianC	C_9	0.8391	0.8382	0.8387	0.8399	0.8393	0.8383	0.8387	0.8389	0.8383	0.8378	0.8368
	C_{10}	0.0376	0.0394	0.0409	0.0402	0.0413	0.0407	0.0399	0.0387	0.0387	0.0404	0.0398
	C_{11}	0.8072	0.8083	0.8114	0.8145	0.8164	0.8176	0.8193	0.8203	0.8190	0.8182	0.8175
	C_{12}	0.0618	0.0622	0.0615	0.0596	0.0588	0.0579	0.0548	0.0514	0.0507	0.0502	0.0488
H-Statlog	C_{13}	0.7672	0.7662	0.7657	0.7647	0.7637	0.7627	0.7642	0.7613	0.7613	0.7627	0.7623
	C_{14}	0.0169	0.0146	0.0154	0.0154	0.0168	0.0176	0.0166	0.0153	0.0184	0.0172	0.0190
	C_{15}	0.6489	0.6456	0.6422	0.6422	0.6400	0.6378	0.6378	0.6322	0.6300	0.6322	0.6300
	C_{16}	0.0358	0.0344	0.0338	0.0327	0.0344	0.0336	0.0336	0.0297	0.0332	0.0297	0.0320
Ionosphere	C_{17}	0.8977	0.9000	0.9004	0.9030	0.9015	0.9023	0.9008	0.9011	0.9015	0.9019	0.9019
	C_{18}	0.0401	0.0381	0.0371	0.0341	0.0322	0.0307	0.0316	0.0316	0.0315	0.0295	0.0295
	C_{19}	0.7927	0.7927	0.7917	0.7938	0.7865	0.7823	0.7771	0.7771	0.7771	0.7760	0.7760
	C_{20}	0.0622	0.0622	0.0632	0.0606	0.0551	0.0606	0.0660	0.0670	0.0700	0.0687	0.0687
Vote	C_{21}	0.9300	0.9300	0.9297	0.9306	0.9306	0.9306	0.9297	0.9291	0.9287	0.9281	0.9269
	C_{22}	0.0195	0.0195	0.0190	0.0198	0.0203	0.0197	0.0187	0.0187	0.0183	0.0177	0.0184
	C_{23}	0.9033	0.9033	0.9025	0.9033	0.9025	0.9024	0.9000	0.8984	0.8976	0.8960	0.8929
	C_{24}	0.0314	0.0314	0.0297	0.0301	0.0304	0.0302	0.0282	0.0293	0.0288	0.0283	0.0323
Iris	C_{25}	0.9333	0.9393	0.9444	0.9444	0.9462	0.9444	0.9385	0.9342	0.9333	0.9316	0.9350
	C_{26}	0.1219	0.1018	0.0933	0.0933	0.0897	0.0921	0.0890	0.0996	0.0961	0.1013	0.0949
	C_{27}	0.8359	0.8564	0.8718	0.8718	0.8795	0.8744	0.8564	0.8538	0.8538	0.8513	0.8564
	C_{28}	0.2647	0.2027	0.1776	0.1776	0.1587	0.1647	0.1467	0.1405	0.1218	0.1260	0.1257
Balance	C_{29}	0.9056	0.9081	0.9094	0.9105	0.9128	0.9141	0.9169	0.9165	0.9162	0.9160	0.9132
	C_{30}	0.0151	0.0141	0.0145	0.0121	0.0112	0.0098	0.0097	0.0100	0.0132	0.0103	0.0109
	C_{31}	0.2567	0.3500	0.4400	0.5301	0.6172	0.6839	0.7607	0.7979	0.8311	0.8672	0.8768
	C_{32}	0.1870	0.2255	0.2541	0.2523	0.1992	0.1865	0.1617	0.1411	0.1156	0.0666	0.0491
Dermatology	C_{33}	0.7586	0.7590	0.7630	0.7667	0.7678	0.7696	0.7696	0.7689	0.7685	0.7674	0.7626
	C_{34}	0.0943	0.0934	0.0932	0.0934	0.0941	0.0946	0.0939	0.0952	0.0969	0.0970	0.0978
	C_{35}	0.2178	0.2200	0.2322	0.2539	0.2539	0.2767	0.2817	0.2861	0.2906	0.2880	0.3144
	C_{36}	0.2249	0.2257	0.2225	0.2257	0.2257	0.2416	0.2396	0.2426	0.2472	0.2470	0.2379

En la Tabla 2 se muestra la matriz de decisión X del problema de clasificación que estamos considerando para las nueve bases de datos, donde el elemento X_{ij} es el valor del criterio de decisión i para la alternativa j (donde $i = 1, ..., 36$ y $j = 1, ..., 11$). Por ejemplo, en BreastC para el par (C_1, A_2) tendríamos el elemento X_{12} cuyo valor es 0.6662. Los datos mostrados en esta tabla son los obtenidos con los conjuntos de generalización.

A partir de la matriz de decisión X se construirá la matriz de decisión normalizada N dividiendo cada elemento X_{ij} por la norma Euclídea.

$$N_{ij} = \frac{X_{ij}}{\sqrt{\sum_{j=1}^{11} \left(X_{ij}\right)^2}}, i = 1, ..., 36, j = 1, ..., 11$$

Gracias a esta transformación, conseguimos que los criterios pierdan la dimensión y a partir de ello tratemos con valores adimensionales.

Para asignar los pesos de los distintos criterios, se consultó a un experto sobre la importancia relativa de unos criterios frente a otros, obteniendo los resultados mostrados en la Tabla 3 (se consideran los valores iguales para todas las bases de datos).

Tabla 3: Comparaciones binarias de los criterios

	$\widehat{\mu}_C$	$\widehat{\sigma}_C$	$\widehat{\mu}_{MS}$	$\widehat{\sigma}_{MS}$
$\widehat{\mu}_C$	1	7	5	9
$\widehat{\sigma}_C$	1/7	1	1/3	5
$\widehat{\mu}_{MS}$	1/5	3	1	5
$\widehat{\sigma}_{MS}$	1/9	1/5	1/5	1

Puesto que la matriz de comparaciones debe ser positiva y recíproca con unos en la diagonal principal, el experto solamente necesitó proporcionar los valores de los juicios asociados a la matriz triangular superior. Para ayudarle en su juicio, se empleó la correspondencia mostrada en la Tabla 4.

A la vista de los valores proporcionados por el experto, se puede ver que los criterios $\widehat{\mu}_C$ y $\widehat{\mu}_{MS}$ tienen una mayor importancia que $\widehat{\sigma}_C$ y $\widehat{\sigma}_{MS}$. También se observa que las medidas de centralización y dispersión de C tienen más importancia que las de MS, lo que es habitual en *machine learning*.

Tabla 4: Valor numérico de las comparaciones

Relación	Valor
A_i y A_j son igualmente importantes	1
A_i es moderadamente más importante que A_j	3
A_i es más importante que A_j	5
A_i es mucho más importante que A_j	7
A_i es extremadamente más importante que A_j	9

Los valores de escala 2, 4, 6 y 8 representan compromisos entre los valores anteriores

A partir de los valores de la Tabla 3, se calcula el vector de pesos asociados a los criterios (W). Los elementos del vector W se obtienen mediante el cálculo del vector propio de la matriz de comparaciones binarias de los criterios. Tras los cálculos, obtenemos el siguiente vector, $W = \{0{,}6485, 0{,}1100, 0{,}2009, 0{,}0407\}$. Este vector se corresponde con una única base de datos, como nosotros empleamos 9 conjuntos de datos diferentes, necesitamos un vector de 36 elementos. Para conseguirlo, dividimos cada uno de los elementos de W entre 9, asignando a cada uno de los criterios $(\widehat{\mu}_C, \widehat{\sigma}_C, \widehat{\mu}_{MS}, \widehat{\sigma}_{MS})$ el valor que le corresponda en función del vector propio W obtenido de la Tabla 3. Esto es así para que la suma de todos los elementos del vector W de nuestro problema sumen 1 y, por tanto, los pesos estén normalizados. Los valores del vector W se pueden observar en la Tabla 5.

Una vez obtenida la matriz de decisión normalizada (N), debemos tener en cuenta los pesos de los criterios (W) para construir la matriz de decisión normalizada ponderada (V). Para ello, se multiplica cada elemento N_{ij} por el peso del criterio i-ésimo.

$$V_{ij} = W_i N_{ij},\ i = 1, ..., 36, j = 1, ..., 11$$

En este punto, se puede determinar la solución ideal positiva A^+ y la solución ideal negativa A^-. Para la solución ideal positiva $A^+ = \left[A_1^+, ..., A_{36}^+\right]$, donde A_i^+ es el máximo valor de V_{ij}, para $j = 1, ..., 11$, para el caso de los criterios que representan atributos deseables ($\widehat{\mu}_C$ y $\widehat{\mu}_{MS}$) o el mínimo valor de V_{ij} para aquellos que representan atributos no deseables ($\widehat{\sigma}_C$ y $\widehat{\sigma}_{MS}$).

Tabla 5: vector W, matriz de decisión normalizada ponderada y solución ideal positiva y negativa

		W	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A^+	A^-
BreastC	C_1	0.0721	0.0330	0.0330	0.0330	0.0328	0.0326	0.0327	0.0328	0.0328	0.0322	0.0321	0.0315	0.0330	0.0315
	C_2	0.0122	0.0056	0.0055	0.0058	0.0053	0.0050	0.0052	0.0046	0.0049	0.0059	0.0063	0.0064	0.0046	0.0064
	C_3	0.0223	0.0086	0.0087	0.0088	0.0089	0.0090	0.0094	0.0097	0.0102	0.0112	0.0125	0.0127	0.0127	0.0086
	C_4	0.0045	0.0022	0.0023	0.0022	0.0022	0.0023	0.0022	0.0022	0.0021	0.0016	0.0014	0.0013	0.0013	0.0023
Pima	C_5	0.0721	0.0328	0.0328	0.0328	0.0327	0.0327	0.0326	0.0326	0.0325	0.0325	0.0323	0.0323	0.0328	0.0323
	C_6	0.0122	0.0057	0.0057	0.0055	0.0052	0.0052	0.0055	0.0051	0.0054	0.0057	0.0059	0.0058	0.0051	0.0059
	C_7	0.0223	0.0092	0.0094	0.0096	0.0098	0.0099	0.0101	0.0103	0.0105	0.0106	0.0107	0.0108	0.0108	0.0092
	C_8	0.0045	0.0026	0.0025	0.0024	0.0022	0.0019	0.0017	0.0017	0.0018	0.0018	0.0019	0.0018	0.0017	0.0026
AustralianC	C_9	0.0721	0.0326	0.0326	0.0326	0.0326	0.0326	0.0326	0.0326	0.0326	0.0326	0.0326	0.0325	0.0326	0.0325
	C_{10}	0.0122	0.0052	0.0055	0.0057	0.0056	0.0057	0.0056	0.0055	0.0054	0.0054	0.0056	0.0055	0.0052	0.0057
	C_{11}	0.0223	0.0100	0.0100	0.0100	0.0101	0.0101	0.0101	0.0101	0.0102	0.0101	0.0101	0.0101	0.0102	0.0100
	C_{12}	0.0045	0.0022	0.0023	0.0022	0.0022	0.0021	0.0021	0.0020	0.0019	0.0018	0.0018	0.0018	0.0018	0.0023
H-Statlog	C_{13}	0.0721	0.0327	0.0326	0.0326	0.0326	0.0325	0.0325	0.0326	0.0324	0.0324	0.0325	0.0325	0.0327	0.0324
	C_{14}	0.0122	0.0055	0.0048	0.0050	0.0051	0.0055	0.0058	0.0054	0.0050	0.0060	0.0056	0.0062	0.0048	0.0062
	C_{15}	0.0223	0.0102	0.0102	0.0101	0.0101	0.0101	0.0100	0.0100	0.0100	0.0099	0.0100	0.0099	0.0102	0.0099
	C_{16}	0.0045	0.0022	0.0021	0.0020	0.0020	0.0021	0.0020	0.0020	0.0018	0.0020	0.0018	0.0019	0.0018	0.0022
Ionosphere	C_{17}	0.0721	0.0324	0.0325	0.0325	0.0326	0.0326	0.0326	0.0325	0.0325	0.0326	0.0326	0.0326	0.0326	0.0324
	C_{18}	0.0122	0.0066	0.0062	0.0061	0.0056	0.0053	0.0050	0.0052	0.0052	0.0051	0.0048	0.0048	0.0048	0.0066
	C_{19}	0.0223	0.0102	0.0102	0.0102	0.0102	0.0101	0.0100	0.0100	0.0100	0.0100	0.0099	0.0099	0.0102	0.0099
	C_{20}	0.0045	0.0019	0.0019	0.0020	0.0019	0.0017	0.0019	0.0021	0.0021	0.0022	0.0021	0.0021	0.0017	0.0022
Vote	C_{21}	0.0721	0.0326	0.0326	0.0326	0.0326	0.0326	0.0326	0.0326	0.0325	0.0325	0.0325	0.0325	0.0326	0.0325
	C_{22}	0.0122	0.0056	0.0056	0.0055	0.0057	0.0058	0.0056	0.0054	0.0053	0.0052	0.0051	0.0053	0.0051	0.0058
	C_{23}	0.0223	0.0101	0.0101	0.0101	0.0101	0.0101	0.0101	0.0100	0.0100	0.0100	0.0100	0.0100	0.0101	0.0100
	C_{24}	0.0045	0.0021	0.0021	0.0020	0.0020	0.0020	0.0020	0.0019	0.0019	0.0019	0.0019	0.0022	0.0019	0.0022
Iris	C_{25}	0.0721	0.0324	0.0326	0.0328	0.0328	0.0329	0.0328	0.0326	0.0324	0.0324	0.0324	0.0325	0.0329	0.0324
	C_{26}	0.0122	0.0069	0.0057	0.0053	0.0053	0.0051	0.0052	0.0050	0.0056	0.0054	0.0057	0.0053	0.0050	0.0069
	C_{27}	0.0223	0.0098	0.0101	0.0102	0.0102	0.0103	0.0103	0.0101	0.0100	0.0100	0.0100	0.0101	0.0103	0.0098
	C_{28}	0.0045	0.0032	0.0025	0.0022	0.0022	0.0019	0.0020	0.0018	0.0017	0.0015	0.0015	0.0015	0.0015	0.0032
Balance	C_{29}	0.0721	0.0323	0.0324	0.0325	0.0325	0.0326	0.0326	0.0327	0.0327	0.0327	0.0327	0.0326	0.0327	0.0323
	C_{30}	0.0122	0.0069	0.0065	0.0066	0.0055	0.0051	0.0045	0.0045	0.0046	0.0061	0.0047	0.0050	0.0045	0.0069
	C_{31}	0.0223	0.0039	0.0053	0.0066	0.0080	0.0093	0.0103	0.0115	0.0120	0.0125	0.0131	0.0132	0.0132	0.0039
	C_{32}	0.0045	0.0021	0.0026	0.0029	0.0029	0.0023	0.0021	0.0018	0.0016	0.0013	0.0008	0.0006	0.0006	0.0029
Dermatology	C_{33}	0.0721	0.0323	0.0323	0.0325	0.0326	0.0327	0.0328	0.0328	0.0327	0.0327	0.0327	0.0325	0.0328	0.0323
	C_{34}	0.0122	0.0055	0.0054	0.0054	0.0054	0.0055	0.0055	0.0055	0.0055	0.0056	0.0056	0.0057	0.0054	0.0057
	C_{35}	0.0223	0.0082	0.0083	0.0088	0.0096	0.0096	0.0105	0.0107	0.0108	0.0110	0.0109	0.0119	0.0119	0.0082
	C_{36}	0.0045	0.0020	0.0020	0.0019	0.0020	0.0020	0.0021	0.0021	0.0021	0.0022	0.0022	0.0021	0.0019	0.0022

Para la solución ideal negativa se seleccionan los valores mínimos de V_{ij} para los criterios que representan atributos deseables o los valores máximos de V_{ij} para los que representan atributos indeseables.

$$A^+ = \left\{V_1^+, ..., V_{36}^+\right\} = \left\{\left(\max_j V_{ij}\right), \left(\min_j V_{ij}\right)\right\}$$

$$A^- = \left\{V_1^-, ..., V_{36}^-\right\} = \left\{\left(\min_j V_{ij}\right), \left(\max_j V_{ij}\right)\right\}$$

En la Tabla 5 se muestran los valores del vector W, de la matriz de decisión normalizada ponderada y la solución ideal positiva y negativa.

Cada alternativa $A_j (j = 1, .., 11)$, A^+ y A^- pueden ser representadas geométricamente como un punto en un espacio de 36 dimensiones, donde el eje i-ésimo mide el rendimiento ponderado normalizado de esa alternativa de acuerdo con el criterio $C_i (i = 1, ..., 36)$. Por lo tanto, podemos calcular la distancia Euclídea (d_j^+ y d_j^-) de cada alternativa A_j a la solución ideal positiva (A^+) y a la solución ideal negativa (A^-), respectivamente.

$$d_j^+ = \left\{\sum_{i=1}^{36} \left(V_{ij} - V_i^+\right)^2\right\}^{\frac{1}{2}}, j = 1, ..., 11$$

$$d_j^- = \left\{\sum_{i=1}^{36} \left(V_{ij} - V_i^-\right)^2\right\}^{\frac{1}{2}}, j = 1, ..., 11$$

A partir de las distancias de todas las alternativas a la solución ideal positiva y a la solución ideal negativa, se calcula el coeficiente de proximidad de cada una de las alternativas.

$$R_j = \frac{d_j^-}{d_j^+ + d_j^-}; \ j = 1, ..., 11$$

El valor de R_j siempre pertenecerá al intervalo $[0, 1]$. Si $R_j = 1$ ($d_j^+ = 0$), entonces $A_j = A^+$ (solución ideal). Si $R_j = 0$ ($d_j^- = 0$), entonces $A_j = A^-$. Es decir, cuanto más próximo esté el índice de una alternativa a 1, mejor alternativa será. En la Tabla 6 se muestran las distancias de cada alternativa a la solución ideal positiva y negativa y los coeficientes de proximidad.

Tabla 6: Distancias y coeficientes de proximidad

	d^+	d^-	R
A_1	0.0120	0.0022	0.1562
A_2	0.0105	0.0032	0.2339
A_3	0.0093	0.0042	0.3094
A_4	0.0077	0.0056	0.4229
A_5	0.0065	0.0069	0.5159
A_6	0.0053	0.0081	0.6032
A_7	0.0042	0.0094	0.6897
A_8	0.0035	0.0099	0.7372
A_9	0.0036	0.0102	0.7416
A_{10}	0.0029	0.0114	**0.7983**
A_{11}	0.0031	0.0118	0.7909

Para terminar, se ordenan todas las alternativas según su coeficiente de proximidad y se selecciona la alternativa con mayor índice como la mejor de todas (el símbolo “>” significa “mejor que”).

$$A_{10} > A_{11} > A_9 > A_8 > A_7 >$$
$$A_6 > A_5 > A_4 > A_3 > A_2 > A_1$$

Al ser la alternativa A_{10} la que obtiene un mayor coeficiente de proximidad, es la que consideraremos como la mejor alternativa. Esta alternativa se corresponde con un valor de λ =0.1.

6. Conclusiones y trabajos futuros

La selección de un modelo de red neuronal concreto perteneciente al primer frente de Pareto es un problema con el que se enfrentan todos los algoritmos multi-objetivo. Es por esto que proponemos el uso de métodos de decisión para ayudar a resolver este problema.

En este trabajo aplicamos la técnica TOPSIS para determinar cuál es la mejor agregación de los modelos que se encuentran en los extremos del primer frente de Pareto para formar el *ensemble*, ya que pensamos que con una combinación de ambos modelos se obtendrán mejores resultados en generalización.

Tras la aplicación de la técnica TOPSIS, podemos concluir que la mejor alternativa de todas las consideradas es la A_{10}. Esta alternativa es la que se obtiene al utilizar un valor de λ =0.1. Al utilizar dicho factor, el modelo del extremo inferior del primer frente de Pareto tiene una importancia del 90 % y el modelo del extremo superior del 10 %. Por esto, afirmamos que el modelo del extremo inferior del frente es el que tiene mayor importancia ya que presenta una mayor precisión y sensibilidad sobre los conjuntos de generalización de las bases de datos propuestas.

Este estudio sugiere varias líneas de investigación futuras. Primero, nos proponemos analizar como afectan las decisiones tomadas por el experto en el problema. Para ello estudiaremos como se modifican los resultados al variar ligeramente los valores de las preferencias dados por el experto. Otra posibilidad sería consultar a varios expertos y ponderar los valores de sus juicios según su experiencia. Por último, se podría emplear el coeficiente de variación de C y MS en lugar de las medias y las desviaciones típicas como criterios para decidir que alternativa es la que produce mejores resultados.

7. Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el proyecto TIN 2008-06681-C06-03 y TIN 2008-06872-C04-04 (MICYT-FEDER) y por los proyectos de la Junta de Andalucía P08-TIC-3745 y P07-TIC02970. La investigación de M. Cruz-Ramírez ha sido financiada por P08-TIC-3745 de la Junta de Andalucía.

Referencias

[1] Abbass, H.A., Sarker, R., Newton, C.: PDE: a Pareto-frontier differential evolution approach for multi-objective optimization problems. En: Proceedings of the 2001 Congress on Evolutionary Computation. Volumen 2, Seoul, South Korea (2001) 971-978

[2] Asuncion, A., Newman, D.J.: UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences (2007) http://www.ics.uci.edu/ /∼mlearn/MLRepository.html

[3] Fernández, J.C., Hervás, C., Martínez, F.J., Gutiérrez, P.A., Cruz, M.: Memetic pareto differential evolution for designing artificial neural networks in multiclassification problems using cross-entropy versus sensitivity. En: Hybrid Artificial Intelligence Systems. Volumen 5572, Springer Berlin / Heidelberg (2009) 433-441

[4] Fernández, J.C., Martínez, F.J., Hervás, C., Gutiérrez, P.A.: Sensitivity versus accuracy in multi-class problems using memetic pareto evolutionary neural networks. IEEE Transactions Neural Networks, accepted. Doi: 10.1109/TNN.2010.2041468, 2010

[5] García Cascales, M.S., Lamata, M.T.: Multi-criteria analysis for a maintenance management problem in an engine factory: Rational choice. Journal of Intelligent Manufacturing. Doi: 10.1007/s10845-009-0290-x

[6] Gräning, L., Jin, Y., Sendhoff, B.: Generalization improvement in multi-objective learning En: International Joint Conference on Neural Networks, 2006. IJCNN '06. Vancouver, BC, Canada (2006) 4839-4846

[7] Hwang, C., Yoon, K.: Multiple attribute decision making: Methods and application. Springer-Verlag (1981)

[8] Ou, G., Murphey, Y.L.: Multi-class pattern classification using neural networks. Pattern Recognition. Volumen 40 (2007) 4-18

[9] Pádua, A., Takahashi, R., Azevedo, M., Teixeira, R.A.: Multi-Objective Algorithms for Neural Networks Learning. En: Multi-Objective Machine Learning. Volumen 16, Springer Berlin / Heidelberg (2006) 151-171

[10] Storn, R., Price, K.: Differential Evolution. A fast and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization. Volumen 11 (1997) 341-359