

Una metodología de recomendación de algoritmos de asignación de pesos basada en casos

Héctor Núñez¹, Miquel Sànchez-Marrè¹

¹Knowledge Engineering & Machine Learning group, Universitat Politècnica de Catalunya,
Campus Nord-Edifici C5, Jordi Girona 1-3,
08034 Barcelona
{hnunez, miquel}@lsi.upc.es

Abstract. La idea subyacente en la metodología propuesta en este artículo es proporcionar una herramienta de software que permita a los usuarios hacer la selección mas apropiada de un algoritmo de asignación de pesos (feature weighting) para una base de datos específica. La idea principal consiste en implementar un sistema basado en casos, en donde cada caso está formado por la descripción de una base de datos y el conjunto de los tres algoritmos supervisados y los dos algoritmos no supervisados de asignación de pesos que han mostrado el mejor rendimiento en un proceso de generalización. La recomendación final se hace guiada principalmente por medio de la evaluación de la similitud entre las características de las bases de casos y las características de la base de datos presentada como nuevo problema. La metodología propuesta es el resultado del conocimiento conceptual adquirido sobre algunos algoritmos ampliamente conocidos de asignación de pesos. La metodología implementa un proceso de recuperación basado en casos y recomienda los mejores algoritmos de asignación de pesos para una nueva base de datos desconocida, recuperando de entre la base de casos el que sea más similar al presentado como problema que es la caracterización de un nuevo dominio.

Palabras Clave: Meta learning, Feature Weighting, Asignación de pesos, Sistemas de recomendación, CBR.

1 Introducción

En los últimos años, la selección del algoritmo mas adecuado para alguna tarea específica dentro les área de machine learning (meta-learning) se ha convertido en una tarea difícil de resolver. Esto se debe principalmente a la aparición de diversas herramientas de software, las cuales ofrecen una gran variedad de modelos de aprendizaje y algoritmos. El proceso de selección de algoritmos se convierte entonces

en un proceso iterativo. El analista debe seleccionar un conjunto pequeño de posibles soluciones dentro de una amplia gama de algoritmos disponibles. Después de este paso de pre selección, se realiza un proceso de evaluación para tomar la decisión final. Esta selección final debe ser hecha tomando en cuenta las características específicas del problema a resolver, y la selección será válida solo para ese problema, puesto que es bien conocido que no existe un modelo o algoritmo que sea la mejor solución para *todos* los posibles problemas independientemente de sus características particulares.

La tarea de *asignación de pesos* es sumamente importante para la determinación apropiada de la relevancia de los atributos en todos los algoritmos basados en instancias (instance-based). Existe una gran cantidad de algoritmos de asignación de pesos, pero no hay uno que pueda ser considerado como el mejor para todos los posibles dominios o bases de datos. El objetivo de este artículo es proponer una nueva metodología que seleccione el algoritmo/s más apropiado/s para una nueva base de datos específica, por medio de un proceso de razonamiento basado en casos (CBR). La nueva base de datos se toma como el problema a resolver desde el punto de vista de CBR. La recomendación final se hará encontrando de entre la base de casos del sistema CBR el que sea más similar al problema actual. La implementación de la base de casos se hace utilizando una estructura de memoria plana, intentando conseguir una mayor precisión en el proceso de recuperación al evaluar la similitud entre el problema actual y *todos* los casos presentes en la base de casos, esto indica claramente que implementaremos un sistema CBR plano.

1.1 Trabajo Relacionado

En el campo de machine learning, se han utilizado diversas metodologías para resolver el problema de la selección de algoritmos, enfocadas principalmente a la tarea de clasificación. Un ejemplo de estas metodologías es la propuesta por Schaffer en [Schaffer, 1993], que sugiere implementar un proceso de fuerza bruta para seleccionar el algoritmo más apropiado: todos los algoritmos disponibles serán ejecutados para el problema actual y su precisión será evaluada por medio de un proceso de cross-validation, el algoritmo que consiga la máxima puntuación en términos de precisión será el seleccionado como solución al problema.

Otro ejemplo es el proyecto Statlog¹ [Michie *et al.*, 1994], que compara el rendimiento de un conjunto fijo de algoritmos sobre algunas bases de datos, específicamente, veintitrés algoritmos son evaluados sobre veintiuna base de datos. El objetivo de Statlog es la derivación automática de reglas basadas en las características de los datos que guíen la selección del clasificador. Se utilizan técnicas de machine learning en la creación de reglas asociando las características de los datos con el rendimiento de los clasificadores, donde el rendimiento está definido en términos de la precisión de la predicción de etiquetas de clase.

Un sistema similar aplicado a la selección de modelos es el propuesto en [Kohavi *et al.*, 1997], donde se presenta una biblioteca llamada MLC++. La propuesta principal consiste en aplicar todos los algoritmos disponibles al problema actual, y la

¹ see <http://www.liacc.up.pt/ML/statlog>

selección final se hará tomando en cuenta el rendimiento de cada uno de los algoritmos.

Otra solución propuesta para resolver el problema de la selección de clasificadores es la presentada en [Kalousis and Theoharis, 1999] llamada NOEMOS. En NOEMOS, la principal idea es hacer la selección de modelos tomando en cuenta las características de los datos. El punto de partida es definir las características de los datos a tomar en cuenta y utilizarlos para implementar un algoritmo de nearest neighbour. Un inconveniente de esta propuesta es que los resultados obtenidos son tan solo de un 58% de precisión para los algoritmos propuestos.

En [Lindner and Studer, 1999] se propone el hacer la selección de algoritmos en base a un proceso CBR. Los autores presentan una herramienta llamada “Algorithm Selection Tool” (AST) que muestra algunas ventajas con respecto a otros sistemas de recomendación de algoritmos. Una característica importante en AST es que el/la usuario/a del sistema no solo obtiene una recomendación acerca de cual algoritmo debería ser aplicado, sino que adicionalmente obtiene una explicación de por que se le hace una recomendación específica. La explicación ofrecida se muestra en forma de experiencias pasadas registradas en la base de casos. Adicionalmente, es posible agregar nuevos algoritmos sin la necesidad de efectuar pruebas de dicho algoritmo sobre todas las bases de datos consideradas en el sistema.

Los miembros del proyecto METAL han desarrollado una implementación de un proceso de meta-learning para algoritmos de clasificación que se encuentra disponible en una pagina web. Los resultados del proyecto METAL son bastante buenos y puede hacerse uso público del sistema accediendo a la pagina: <http://www.metal-kdd.org>.

Todos los sistemas de recomendación descritos en esta sección se enfocan hacia la selección de algoritmos de clasificación, y hasta donde nuestro conocimiento alcanza, no tenemos noticias de experimentos similares que aborden el problema de la recomendación de algoritmos de asignación de pesos. En este artículo, hacemos la propuesta de una metodología de recomendación de algoritmos de asignación de pesos.

1.2 Organización

El artículo esta estructurado de la siguiente manera. En la sección 2 se describe la arquitectura del sistema CBR de recomendación de algoritmos. La sección 3 muestra los detalles de la estructura de la base de casos y el formalismo de representación de los casos. La sección 4 describe los detalles de la experimentación empírica. En la sección 5 se analiza la evaluación y rendimiento de la metodología propuesta. Finalmente, algunas conclusiones y vertientes para trabajo futuro relacionadas con el tema se presentan en la sección 6.

2 Arquitectura del Sistema de Recomendación de Algoritmos Basado en Casos

Una de las ventajas que se tienen al utilizar un sistema CBR es que se emplean operadores de similitud en lugar de las sumamente rígidas reglas como las derivadas en Statlog. Otra ventaja es que el sistema tiene un mecanismo de mantenimiento de la base de casos muy sencillo, que consiste en simplemente agregar nuevos casos a la base de casos. También es posible implementar un proceso de explicación de resultados con base en las experiencias pasadas disponible en la base de casos que pueden ser muy útiles para los usuarios.

La arquitectura general de la metodología de recomendación propuesta se muestra en la fig. 1. El proceso de la recomendación de algoritmos es una decisión basada principalmente en dos factores:

- Datos del problema
- Experiencia adquirida

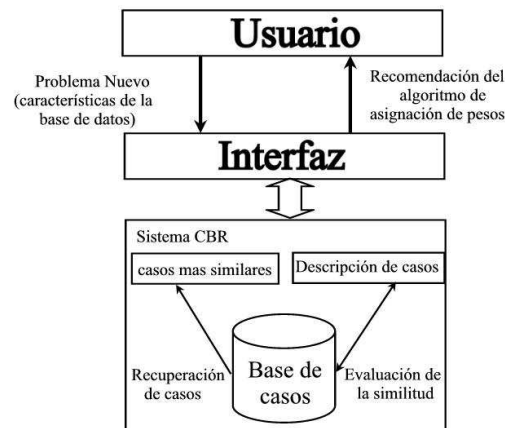


Fig. 1. Arquitectura del sistema CBR de recomendación de algoritmos.

El sistema CBR recibe los datos del problema actual a través de una interfaz de usuario. El flujo general es que el usuario especifica sus requerimientos describiendo las características del problema. Cada caso en la base de casos está definido por su descripción de un problema específico, y la parte de solución, que indica cuales son los cuatro mejores algoritmos para ese problema en particular. De los cuatro algoritmos propuestos como mejor solución, tres de ellos son aplicables a dominios supervisados y otro a dominios no supervisados. La experiencia que se encuentra registrada de manera implícita en cada caso, es la precisión obtenida al aplicar cada uno de los algoritmos registrados como las mejores opciones para la asignación de pesos en un proceso de clasificación. Estos valores de precisión son obtenidos al realizar un proceso de ten-fold cross validation, sobre la base de datos. Para la

evaluación de la similitud se utilizó la medida de distancia *L'Example* [Sánchez-Marrè *et al.* 1998], y cuando fue necesario por las características de los algoritmos y de los datos, se utilizó el método de discretización CAIM descrito en [Kurgan and Cios 2001]. Se determinó el uso de *L'Example* y CAIM por los buenos resultados obtenidos después de exhaustivos experimentos realizados con otras medidas de similitud y otros métodos de discretización descritos en [Núñez *et al.*, 2003]. Es importante destacar que es frecuente encontrar definiciones de nuevos problemas con descripciones de bases de datos incompletas, en estos casos, los atributos que no hayan sido descritos se tratarán como valores desconocidos (missing values).

3 Estructura de la Base de Casos

Puesto que un sistema CBR plano se basa principalmente en la evaluación de la similitud y la relevancia de los atributos, y la mayoría de los métodos de asignación de pesos y medidas de similitud son aplicables solo a entornos supervisados, es necesario agrupar los casos y asignar a cada grupo resultante una etiqueta de clase. Para la propuesta de la metodología hemos experimentado con dos diferentes métodos de agrupamiento.

El primer método, simplemente asigna como etiqueta de clase para cada caso el nombre del algoritmo que presenta el mejor rendimiento sobre la base de datos descrita en ese caso. La cantidad de posibles etiquetas de clase es igual al total de diferentes algoritmos probados. En nuestro caso, se han realizado pruebas con trece diferentes algoritmos de asignación de pesos, por lo que se obtienen trece diferentes etiquetas de clase que son una para cada algoritmo: sin pesos (NW), Ganancia de información (IG), extensión de Daelemans and Van Den Bosch para ganancia de información (IGDB), proyección de atributos (PROJ), RELIEF (RELF), asignación global de pesos por distribución de clases (CDWG), distribución de clases (CD), distribución de valores (VD), distribución de clases-valores (CVD), asignación local por distribución de clases (CDWL), Value-Difference Metric (VDM), asignación de pesos locales basados en entropía (EBL), gradiente descendiente (GD), método 1 no supervisado basado en entropía (UEB-1) y método 2 no supervisado basado en entropía (UEB-2).

El segundo método implementado para la asignación de etiqueta de clase para cada caso asigna como nombre de etiqueta la concatenación de los nombres de los dos mejores algoritmos para cada caso. De acuerdo a la base de casos que se tiene hasta el momento, se obtuvieron quince diferentes etiquetas de clase. Las etiquetas de clase obtenidas por los métodos se muestran en la tabla 2.

Como se ha mencionado antes, cada caso en la base de casos contiene información acerca de una base de datos específica y el rendimiento de los trece algoritmos probados sobre esa base de datos. Las características de las bases de datos tomadas en cuenta para la descripción de los casos, contienen valores que pueden ser calculados para toda la base de datos o bien para algún conjunto particular de atributos de acuerdo a su tipo.

- Número de instancias (N_i)

- Número de atributos (Na)
- Número de atributos continuos (Ca)
- Número de atributos discretos ordenados (Doa)
- Número de atributos discretos no ordenados (Dno)
- Número de clases (Nc)
- Porcentaje de valores desconocidos (Mp)
- Promedio del índice de correlación entre atributos (Ci)

La idea principal es proporcionar al sistema toda la información disponible acerca de la base de datos en general y con respecto a los atributos en particular.

La solución para cada caso esta representada por los nombres de los cuatro algoritmos de asignación de pesos con los que se obtiene el mejor rendimiento. Para cada caso, se almacena el nombre de los tres mejores algoritmos supervisados y el mejor algoritmo no supervisado, la precisión obtenida por cada uno de estos algoritmos forma también parte de la solución. En la fig. 2 se muestra una representación de alto nivel de los casos.

Descripción de la base de datos							Cuatro mejores algoritmos y su precisión								Etiqueta de clase	
Ni	Na	Ca	Doa	Dno	Nc	Mp	Ci	B1	A1	B2	A2	B3	A3	B4	A4	Clase

Fig. 2. Representación de casos en el sistema.

4 Experimentos Sobre la Calidad de la Recomendación

Para evaluar el rendimiento de nuestro sistema de recomendación, implementamos un proceso de five-fold cross validation. Decidimos utilizar five-fold cross validation en lugar del mas ampliamente utilizado ten-fold cross validation, debido a la cantidad reducida de casos que conforman nuestra base de casos. Las pruebas se realizaron con una nueva base de datos que contiene información acerca del rendimiento de trece algoritmos de asignación de pesos sobre treinta y cuatro diferentes bases de datos, de las que 33 fueron tomadas del repositorio de la UCI [Blake and Merz, 1998] y una con información de la calidad del aire de la ciudad de México. Las características de las treinta y cuatro bases de datos se muestran en la tabla 1.

La base de la evaluación de la metodología es el conocimiento obtenido a partir de la tabla 2. En esta tabla se muestran los cuatro mejores algoritmos (en términos de precisión en la clasificación) para cada base de datos. Se tomaron en cuenta los tres mejores algoritmos supervisados y el mejor algoritmo no supervisado. Para medir la precisión en el proceso de five-fold cross validation se consideraron cuatro diferentes criterios:

- En el primer criterio, se considera que se ha hecho una buena recomendación si el algoritmo recomendado en primer lugar para el caso actual es el mismo que el algoritmo presente como mejor resultado para el caso recuperado de la base de casos.

- En el segundo criterio, el sistema considera que se ha hecho una buena recomendación si el mejor algoritmo del caso recuperado es uno de los tres mejores métodos para el caso actual.
- El tercer criterio asume que una recomendación es acertada si una de las dos mejores opciones del caso recuperado está en el conjunto de los tres mejores casos del caso actual.
- Finalmente, el cuarto criterio y el mas relajado en cuanto a exigencia, considera que se hace una buena recomendación si uno de los tres algoritmos del caso recuperado está entre las tres mejores opciones del caso actual.

Base de datos	Nombre corto	Número de instan.	#Cont.	Discret. Ord.	Discret. No ord.	#Clases	%Mis	Promedio de correl.
Air pollution	AP	365	5	0	0	4	0	0.2459
Annealing	AN	798	6	29	3	6	64.9%	0.0134
Audiology	AD	200	0	8	61	24	0	0
Australian	AS	690	6	4	4	2	0	0.0999
Auto	AU	205	15	0	8	7	0.004	0.1029
Bands	BA	512	20	0	20	2	4.87	0.0732
Breast Cancer	BC	699	0	9	0	2	0	0.5934
Bridges	BR	108	3	0	8	3	0.06	0.1592
Cleveland	CL	303	5	2	6	2	0	0.0251
Contraceptive	CM	1473	2	4	3	3	0	0.3103
Credit	CR	690	6	0	9	2	0.64%	0.0975
Dermatology	DE	366	1	0	33	6	0	0
Ecoli	EC	336	7	0	0	8	0	0.1571
Flag	FL	194	3	7	18	8	0	0.0605
German	GE	1000	7	0	13	2	0	0.0631
Glass	GL	214	9	0	0	7	0	0.0915
Hayes-Roth	HR	132	0	0	4	3	0	0
Hepatitis	HE	155	6	0	13	2	5.7	0.1015
Horse-Colic	HC	301	7	0	16	2	30	0
Ionosphere	IO	351	34	0	0	2	0	0.1179
Iris	IR	150	4	0	0	3	0	0.2977
LED	led	300	0	0	7	10	0	0
LED-17	LI	200	0	0	24	10	0	0
Liver Disorders	LD	345	6	0	0	2	0	0.2554
Machine	MA	209	6	0	0	8	0	0.249
Pima Indians	PI	768	8	0	0	2	0	0.1471
Post-Operative	PO	90	1	7	0	3	0	0
Soybean (large)	SL	307	0	6	29	19	21.7	0
Vehicle	VE	946	18	0	0	4	0	0.2379
Votes	VO	435	0	0	16	2	7.3	0
Waveform	WF	300	21	0	0	3	0	0.0246
Waveform-40	WA	300	40	0	0	3	0	0.0069

Wine	WI	178	13	0	0	3	0	0.0849
Zoo	ZO	90	0	0	16	7	0	0
Hayes-Roth15D	HR15D	132	0	0	19	3	0	0
Iris15C	IR15C	150	19	0	0	3	0	0.0053

Tabla 1. Características de las bases de datos involucradas en los experimentos e determinación de la calidad de la recomendación de la metodología propuesta.

Para hacer la recuperación en la base de casos con información de las bases de datos, se utilizaron diversas medidas de similitud [Núñez *et al.*, 2003] y diversos algoritmos de asignación de pesos que han mostrado buenos resultados [Núñez *et al.*, 2003a]. Todas las pruebas fueron evaluadas tomando en cuenta los cuatro criterios descritos y los dos diferentes métodos de agrupamiento (clustering). Obviamente, la precisión en la generalización es mas alta cuando se utiliza el cuarto criterio que es el mas relajado.

	1 st mejor Algoritmo/ Precisión	2 nd mejor Algoritmo/ Precisión	3 rd mejor Algoritmo/ Precisión	1 st mejor Algoritmo no sup./ Precisión	Class(1)	Class(2)
AP	CVD/100	PROJ/99.75	IG/99.47	GD/91.70	CVD	CVDPR
AN	EBL/94.25	CD/94	VDM/93.28	GD/91.97	EBL	EBLCD
AD	IG/80	CVD/78	CD/78	UEB-1/76.50	IG	IGCVD
AS	PROJ/83.19	IG/82.90	RELF/82.90	GD/82.61	PRO	PROJI
AU	PROJ/84.72	VDM/80.83	IG/80.79	GD/76.79	PRO	PROJV
BA	VDM/81.49	EBL/80.37	CVD/78.70	UEB-1/73.70	VD	VDME
BC	EBL/96.70	VDM/96.32	PROJ/96.10	GD/94.40	EBL	EBLVD
BR	IG/94.79	IGDB/94.79	CVD/94.03	UEB-1/90.13	IG	IGIGD
CL	IGDB/78.51	GD/78.46	VDM/77.81	GD/78.46	IGD	IGDBG
CM	VDM/46.71	EBL/45.89	VD/45.89	UEB-1/44.81	VD	VDME
CR	IG/82.75	RELF/82.61	EBL/82.17	UEB-2/81.45	IG	IGREL
DE	CD/97.05	VD/96.22	CVD/95.85	GD/92.05	CD	CDVD
EC	PROJ/83.26	VDM/81.75	GD/81.66	GD/81.66	PRO	PROJV
FL	IGDB/65.47	VDM/64.83	EBL/64.30	GD/53.44	IGD	IGDBV
GE	VDM/72.50	EBL/70.80	PROJ/71.10	UEB-2/69.80	VD	VDME
GL	IG/80.96	VDM/77.27	CVD/77.15	GD/73.52	IG	IGVD
HR	EBL/85.03	VDM/85.03	CDWL/78.2	UEB-1/76.56	EBL	EBLVD
HE	EBL/84.46	GD/82.20	VDM/81.35	UEB-1/82.20	EBL	EBLGD
HC	VD/80.63	EBL/79.90	IG/79.36	UEB-1/73.60	VD	VDEBL
IO	EBL/94.87	VDM/93.44	CVD/92	GD/90.86	EBL	EBLVD
IR	EBL/96	CVD/96	IG/94.67	UEB-2/94	EBL	EBLCV
Led	IG/68	RELF/68	CVD/68	GD/67.67	IG	IGREL
LI	RELF/64	IG/62.50	CD/60.50	UEB-2/38	REL	RELFI
LD	EBL/69.85	VDM/69.24	CVD/65.56	UEB-2/63.54	EBL	EBLVD
MA	EBL/72.68	VDM/72.58	IGDB/72.05	UEB-2/69.67	EBL	EBLVD
PI	VDM/73.17	EBL/72.67	PROJ/71.19	GD/70.47	VD	VDME
PO	VD/66.67	EBL/63.33	VDM/62.22	UEB-1/56.67	VD	VDEBL
SL	PROJ/93.25	EBL/92.97	VDM/92.97	GD/90.04	PRO	PROJE
VE	IGDB/70.21	VDM/69.60	EBL/69.50	GD/66.74	IGD	IGDBV
VO	PROJ/97.04	VDM/96.74	CD/96.10	GD/94.63	PRO	PROJV
WF	IG/76.88	PROJ/76.17	GD/75.82	GD/75.82	IG	IGPRO
WA	EBL/81.67	VDM/81.33	IG/81.33	GD/73	EBL	EBLVD
WI	VDM/99.52	IGDB/97.98	EBL/97.87	GD/95.74	VD	VDMI
ZO	PROJ/98.09	VD/98	IGDB/97.09	GD/96.09	PRO	PROJV
HR15D	EBL/76.65	CD/69.01	VDM/68.35	UEB-2/45.33	EBL	EBLCD

	1 st mejor Algoritmo/ Precisión	2 nd mejor Algoritmo/ Precisión	3 rd mejor Algoritmo/ Precisión	1 st mejor Algoritmo no sup./ Precisión	Class(1)	Class(2)
IR15C	VD/96.67	CD/96	FBI/95.33	UEB-2/95.33	VD	VDCD

Tabla 2. Algoritmos que presentan el mejor rendimiento para cada base de datos y etiquetas de clase obtenidas después de aplicar los dos métodos de agrupamiento descriptos.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NW	9.04	20.96	14.76	9.04	9.04	2.86	9.04	90.96	90.96
IG	11.9	23.8	17.62	11.9	11.42	n/a	n/a	n/a	n/a
IGDB	9.04	18.1	14.76	9.04	11.9	n/a	n/a	n/a	n/a
PROJ	9.04	20.96	14.76	9.04	5.72	n/a	n/a	n/a	n/a
RELF	11.9	18.1	14.76	11.9	14.76	n/a	n/a	n/a	n/a
CDWG	6.2	15.24	9.04	6.2	2.86	n/a	n/a	n/a	n/a
CD	9.04	20.96	14.76	9.04	9.04	n/a	n/a	n/a	n/a
VD	11.9	20.96	14.76	11.9	11.42	n/a	n/a	n/a	n/a
CVD	11.9	20.96	17.62	11.9	17.62	n/a	n/a	n/a	n/a
CDWL	8.58	23.8	17.62	8.58	8.58	n/a	n/a	n/a	n/a
VDM	9.04	14.76	17.62	9.04	29.04	n/a	n/a	n/a	n/a
EBL	9.04	14.76	17.62	9.04	29.04	n/a	n/a	n/a	n/a
GD	11.9	11.42	15.24	11.9	11.9	n/a	n/a	n/a	n/a
UEB-1	11.9	8.58	15.24	11.9	18.1	n/a	n/a	n/a	n/a
UEB-2	12.38	14.76	15.24	12.38	15.24	n/a	n/a	n/a	n/a
NW	29.04	61.9	31.9	29.04	29.04	35.24	38.1	96.66	93.8
IG	37.62	64.76	37.62	37.62	46.2	n/a	n/a	n/a	n/a
IGDB	34.76	56.2	34.76	34.76	43.34	n/a	n/a	n/a	n/a
PROJ	37.62	58.58	37.62	40.48	40.48	n/a	n/a	n/a	n/a
RELF	31.9	59.04	31.9	31.9	29.04	n/a	n/a	n/a	n/a
CDWG	26.2	53.34	29.04	26.2	29.04	n/a	n/a	n/a	n/a
CD	29.04	61.9	34.76	29.04	29.04	n/a	n/a	n/a	n/a
VD	37.62	61.9	31.9	34.76	31.9	n/a	n/a	n/a	n/a
CVD	37.62	58.58	37.62	34.76	38.1	n/a	n/a	n/a	n/a
CDWL	37.62	61.9	40.48	37.62	34.76	n/a	n/a	n/a	n/a
VDM	31.9	47.14	37.62	31.9	52.38	n/a	n/a	n/a	n/a
EBL	31.9	43.8	37.62	31.9	49.52	n/a	n/a	n/a	n/a
GD	37.62	52.86	43.8	37.62	31.9	n/a	n/a	n/a	n/a
UEB-1	46.66	55.24	49.52	46.66	50.48	n/a	n/a	n/a	n/a
UEB-2	26.66	47.14	26.66	26.66	29.52	n/a	n/a	n/a	n/a
NW	51.9	73.34	51.9	54.76	54.76	52.38	52.38	96.66	93.8
IG	54.76	76.2	54.76	54.76	60.48	n/a	n/a	n/a	n/a
IGDB	57.62	70.48	57.62	57.62	60.48	n/a	n/a	n/a	n/a
PROJ	60.48	70.48	58.1	60.48	57.62	n/a	n/a	n/a	n/a
RELF	54.76	70.48	58.1	54.76	46.2	n/a	n/a	n/a	n/a
CDWG	54.76	64.76	55.24	51.9	54.76	n/a	n/a	n/a	n/a
CD	54.76	73.34	54.76	54.76	54.76	n/a	n/a	n/a	n/a
VD	51.9	73.34	55.24	54.76	49.04	n/a	n/a	n/a	n/a
CVD	49.04	73.34	54.76	51.9	61.42	n/a	n/a	n/a	n/a
CDWL	57.62	70.48	60.48	57.62	54.76	n/a	n/a	n/a	n/a
VDM	49.04	55.72	63.8	49.04	63.8	n/a	n/a	n/a	n/a

FRI	49.04	55.77	63.8	49.04	60.96	n/a	n/a	n/a	n/a
GD	57.62	70	66.66	57.62	57.62	n/a	n/a	n/a	n/a
UEB-1	60.96	78.58	67.14	60.96	62.38	n/a	n/a	n/a	n/a
UEB-2	50	52.86	44.28	52.86	58.58	n/a	n/a	n/a	n/a
NW	70	82.86	72.86	70	70	73.33	70	100	93.81
IG	67.14	85.71	70	67.14	70	n/a	n/a	n/a	n/a
IGDB	70	80	72.86	70	72.86	n/a	n/a	n/a	n/a
PROJ	72.86	80	73.33	72.86	72.86	n/a	n/a	n/a	n/a
RELF	70	77.14	79.05	70	61.43	n/a	n/a	n/a	n/a
CDWG	72.86	74.29	76.19	70	72.86	n/a	n/a	n/a	n/a
CD	70	82.86	72.86	70	70	n/a	n/a	n/a	n/a
VD	70	82.86	76.19	70	64.29	n/a	n/a	n/a	n/a
CVD	64.29	82.86	72.86	67.14	85.24	n/a	n/a	n/a	n/a
CDWL	70	85.71	72.86	70	70	n/a	n/a	n/a	n/a
VDM	67.14	68.1	79.05	67.14	84.76	n/a	n/a	n/a	n/a
EBL	67.14	70.95	79.05	67.14	78.57	n/a	n/a	n/a	n/a
GD	70	76.67	75.71	70	75.71	n/a	n/a	n/a	n/a
UEB-1	72.86	88.1	73.33	72.86	74.29	n/a	n/a	n/a	n/a
UEB-2	65.24	74.29	59.52	65.24	70.95	n/a	n/a	n/a	n/a

Tabla 3. Precisión obtenida en el proceso de generalización utilizando Five-fold cross validation sobre la base de datos de meta-learning usando los cuatro criterios descritos, el primer método de agrupamiento descrito y nueve diferentes medidas de similitud: (1) HVDM, (2) IVDM, (3) Euclidea, (4) Manhattan, (5) *L'Example*, (6) Clark, (7) Canberra, (8) SF y (9) MRM.

5 Evaluación

Los resultados de los experimentos realizados están detallados en la tabla 3. Una característica de estos resultados es que para los cuatro criterios usados y para ambos métodos de agrupamiento, las medidas de similitud SF y MRM son las que muestran los mejores resultados. De acuerdo a estos datos, parece ser que el sistema de recomendación muestra un buen rendimiento y puede ser muy útil en la recomendación de algoritmos de asignación de pesos, proporcionando al usuario un conjunto restringido de tres opciones en lugar de buscar el mejor algoritmo entre una gran cantidad de algoritmos que puede representar un coste computacional alto.

Debido a los resultados obtenidos con las diversas medidas de similitud y métodos de agrupamiento, decidimos que nuestro sistema de recomendación estará basado en la medida SF y el primer método de agrupamiento descrito. Puesto que SF no usa pesos en forma explícita, los atributos que describen los casos tendrán todos el mismo peso. En la tabla 3 se indica con n/a los casos en los que se utiliza una medida de similitud que no utiliza pesos, por lo que solo aparece algún valor en la primera fila.

Los resultados del proceso de five-fold cross validation muestran que la precisión en la recomendación que hace nuestro sistema varía desde 90.96% para el criterio más estricto, hasta el 100% para el criterio más relajado. Esto significa que en el 90.96% de las veces, el método recomendado en primer lugar por nuestro sistema es el mismo que el mejor método para el caso actual en el conjunto de entrenamiento. Consideramos que el 90.96% es un valor bastante alto que representa un rendimiento muy aceptable del sistema.

Para confirmar los resultados iniciales obtenidos, se han realizado también pruebas adicionales utilizando otras siete diferentes bases de datos que no estaban presentes en el conjunto de prueba inicial. Para cada una de las siete bases de datos, el sistema recupera el caso más similar. Los resultados de este proceso de recuperación se

muestran en la tabla 4, la cual detalla el rendimiento en términos de porcentaje de precisión obtenidos en un proceso de ten-fold cross validation realizado en forma independiente al sistema. Las últimas tres columnas muestran los tres algoritmos recomendados por el sistema. Como puede observarse, únicamente en una base de datos (Monk-2), los algoritmos recomendados no están dentro de las tres mejores opciones para esa base de datos. Adicionalmente, en cuatro de las siete recomendaciones, se incluye el mejor algoritmo para la base de datos evaluada.

Base de datos	Tres mejores algoritmos			Algoritmos recomendados		
Balance	CDWL/82.14	EBL/69.01	CD/57.43	EBL/69.01	VDM/61	CDWL/82.14
Image	EBL/89.52	RELF/89.52	PROJ/89.52	RELF/89.52	IG/87.62	CD/87.14
Lenses	CVD/87.5	CD/87.5	RELF/87.5	CVD/87.5	PROJ/85	IG/87.45
Monk-2	EBL/47.66	VDM/47.66	CD/47.43	CVD/35.85	PROJ/42.56	IG/35.85
Monk-3	EBL/100	CD/100	IG/100	CVD/92.61	PROJ/98.39	IG/100
Tic-Tac-Toe	EBL/87.79	CVD/87.76	CD/87.76	RELF/79.86	IG/87.76	CD/87.76
WWTP	CVD/88.02	RELF/87.28	CDWL/87.28	CVD/88.02	PRO/85.14	IG/82.47

Tabla 4. Resultados obtenidos para siete bases de datos diferentes usando la medida de similitud SF.

De acuerdo a los resultados de la tabla 4, cuando el sistema usa el cuarto criterio, el porcentaje de recuperaciones exitosas es de 85.71% (seis de siete). Si el tercer criterio se utiliza, el porcentaje de aciertos es de 57.14% (cuatro de siete). Con el segundo criterio, el porcentaje de éxitos es de 42.85% (tres de siete), y finalmente, si se considera el criterio más estricto, el porcentaje de aciertos es de 28.57% (dos de siete).

6 Conclusiones y Trabajo Futuro

Estos resultados nos permiten confirmar que los usuarios de nuestro sistema de recomendación, se verán beneficiados al tener un conjunto de únicamente *tres* posibles algoritmos a probar con un alto grado de certidumbre de que entre estos tres se encuentra el algoritmo más conveniente para su base de datos.

A pesar de que la precisión obtenida para las nuevas bases de datos disminuye un poco con respecto a los resultados del five-fold cross validation, no podemos decir que esto es una influencia negativa en la evaluación de nuestra metodología, y por otro lado, esto es un indicador de que estas bases de datos pueden representar a otros dominios que no habían sido considerados en la base de casos inicial y que pueden colaborar para obtener un mejor rendimiento del sistema en el futuro. De cualquier manera, podemos decir que con el cuarto criterio, el rendimiento del sistema, por ahora, varía desde el 85% al 90%, lo cual es bastante bueno y puede ahorrar a los usuarios mucho tiempo en su búsqueda de un método de asignación de pesos para su base de datos específica.

Esta metodología es sumamente flexible, adaptable y gran capacidad para aprender de nuevas experiencias para ser más competente en el futuro. Mientras más bases de datos se incorporen a la base de casos del sistema, la respuesta será más precisa.

Las principales vertientes que se vislumbran como trabajo futuro, consideran el agregar tantas bases de datos como sea posible para incrementar el rendimiento del sistema y por supuesto el realizar más pruebas de evaluación para confirmar los resultados tan satisfactorios que se han obtenido hasta ahora.

Referencias

- [Blake and Merz, 1998] C.L. Blake, and C.J. Merz. UCI Repository of machine learning databases [<http://www.ics.uci.edu/~mlearn/MLRepository.html>]. Irvine, CA: University of California, Department of Information and Computer Science. 1998.
- [Hilario and Kalousis, 1999] M. Hilario and A. Kalousis. Characterizing Learning Models and Algorithms for Classification. *Technical Report UNIGE-AI-99-01*, CUI-University of Geneva, 1999.
- [Kalousis and Hilario, 2001], A. Kalousis and M. Hilario. Feature Selection for Meta-learning. *In proceedings of the 5th PAKDD*. Hong Kong, April 2001.
- [Kalousis and Theoharis, 1999], A. Kalousis and T. Theoharis. NOEMON: An intelligent Assistant for Classifier Selection. *In Workshop proceedings of the ICML 1999, Workshop 1*.
- [Kohavi *et al.*, 1997] R. Kohavi, P. Langley, and Y. Yun. The utility of feature weighting in nearest-neighbor algorithms. *In Proceedings of the European Conference on Machine Learning (ECML97)*, 1997.
- [Kurgan and Cios, 2001] L. Kurgan and K. J. Cios. Discretisation Algorithm that Uses Class-Attribute Interdependence Maximisation, *Proceedings of the 2001 International Conference on Artificial Intelligence (IC-AI 2001)*, pp.980-987, Las Vegas, Nevada.
- [Linder and Studer, 1999] G. Linder and R. Studer. AST: Support for Algorithm Selection with a CBR Approach. *In C. Giraud-Carrier and B. Pfahringer, editors, Recent Advances in Meta-Learning and Future Work*. Pp 38-47. J. Stefan Institute, 1999.
- [Michie *et al.*, 1994] D. Michie, D.J. Spiegelhalter, C.C. Taylor. *Machine Learning Neural and Statistical Classification*. Ellis Horwood, 1994.
- [Núñez *et al.*, 2003] H. Núñez, M. Sánchez-Marrè and U. Cortés. Similarity Measures in Instance-Based Reasoning. Submitted to *Artificial Intelligence*, 2003.
- [Núñez *et al.*, 2003a] H. Núñez, M. Sánchez-Marrè y U. Cortés. Improving Similarity Assessment with Entropy-Based Local Weighting. *In proceedings of the 5th International Conference on Case-Based Reasoning*, pp. 377-391. ISSN 0302-9743. Trondheim, Norway, June 23 to 26, 2003.
- [Sánchez-Marrè *et al.*, 1998] M. Sánchez-Marrè, U. Cortés, I. R-Roda, and M. Poch. L'Example distance: a new similarity measure for case retrieval. *Procc. of 1st Catalan Conference on Artificial Intelligence (CCIA '98)*, ACIA bulletin 14-15 pp. 246-253. Tarragona, Catalonia, EU.
- [Schaffer, 1993], C. Schaffer. Selecting a classification method by cross-validation. *Machine Learning*. 13:135-143, 1993.
- [Todorovski *et al.*, 2000] L. Todorovski, P. Brazdil, and C. Soares. Report on the experiments with feature selection in meta-level learning. *In Proceedings of the 4th European Conference on Principles on Data Mining and Knowledge Discovery, Workshop on data mining, Decision Support, Meta-Learning and ILP*, pp 27-39, 2000.