El Método Mimético, una Alternativa para la Comprensibilidad de Modelos de "Caja Negra"[1]

Ricardo Blanco-Vega, José Hernández-Orallo, María José Ramírez-Quintana

Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia, C. de Vera s/n, 46022 Valencia, España

{rblanco, jorallo, mramirez}@dsic.upv.es

Resumen. En este artículo estudiamos las propiedades del método mimético, un método general y simple que se emplea para obtener un clasificador similar a cualquier otro clasificador el cual es considerado como un oráculo. El objetivo principal del método es encontrar un clasificador que imite el comportamiento del oráculo, de tal suerte que la fidelidad del clasificador sea alta y tal que el conjunto de reglas generadas permita encontrar un entendimiento y explicación del comportamiento de dicho oráculo. En este artículo resumimos los trabajos experimentales y teóricos realizados por el grupo de investigación MIP[2] en torno al método mimético.

1 Introducción

Dentro del aprendizaje automático existen diversas técnicas, como las redes neuronales, los métodos combinados, etc. que generan modelos que son considerados como cajas negras, dado que transforman las entradas en las salidas sin que sea posible saber cómo esta transformación se lleva a cabo. Desde el punto de vista de la extracción de conocimiento, aunque estos modelos tengan buenas prestaciones tienen el inconveniente que se puede extraer de ellos poco conocimiento útil para poder ser usado en la toma de decisiones, ya que no son comprensibles. Sin embargo, para muchas aplicaciones la comprensibilidad del modelo es crucial. Es por ello que se han propuesto diversos métodos para convertir un modelo incomprensible en una representación simple y comprensible: un conjunto de reglas.

La extracción de reglas de la forma *if-then* es usualmente aceptada como la mejor forma de expresar el conocimiento representado en una caja negra. No porque sea un trabajo fácil, sino porque al final las reglas creadas son más comprensibles para los humanos que otra representación. La mayoría de estas aproximaciones se han centrado en la extracción de reglas a partir de redes neuronales [3][4][10][13][16].

[1] Este trabajo ha sido pracialmente subvencionado por el proyecto ICT for EU-India Cross Cultural Dissemination Project ALA/95/23/2003/077-054, CICYT TIN 2004-7943-C04-02, ANUIES(SUPERA becario No. 5177) y licencia beca comisión de la Dirección General de Institutos Tecnológicos.

[2] http://www.dsic.upv.es/~flip/

Recientemente, en [5][6][7] se ha propuesto una nueva forma de extraer un modelo comprensible a partir de cualquier modelo de caja negra. La idea básica consiste en considerar el modelo inicial como un oráculo y utilizarlo para etiquetar un conjunto de ejemplos generados al azar para entrenar un segundo modelo comprensible. Llamamos a esta técnica mimética ya que el objetivo es que el segundo modelo imite lo máximo posible al oráculo.

Aunque, tal y como mostramos en la siguiente sección, se han analizado varias cuestiones del método mimético, como la precisión del modelo y la utilidad de usar el conjunto de entrenamiento original para generar el modelo mimético, existen otros aspectos que pueden influir en el comportamiento del método que no han sido estudiados:

- El método no se ha aplicado a oráculos que no sean combinación de clasificadores (boosting y bagging), por ejemplo redes neuronales.
- La relación entre la comprensibilidad resultante del modelo y de algunos factores como el grado de poda o el número de los ejemplos generados al azar

En este trabajo analizamos el método con respecto a estas cuestiones, concentrándonos especialmente en cómo obtener un conjunto de reglas corto sin sacrificar demasiado la fidelidad del modelo mimético con respecto al oráculo.

Para ello, definimos una métrica de "calidad" (Q) para que nos sirva de referencia, la cual representa la relación entre la precisión (Acc) y la comprensibilidad (representada simplemente por el número de reglas del modelo mimético):

$$Q = \frac{(\,Acc(Mim) - Acc(Ref)\,) / Acc(Ref)}{(Reglas(Mim) - Reglas(Ref)) / Reglas(Ref)} \qquad (1)$$

El "modelo de referencia" (*Ref*) representa un modelo comprensible aprendido directamente con el conjunto de datos original, como C4.5, mientras que Mim representa el modelo mimético (posiblemente también C4.5). Obviamente, si los resultados con el procedimiento mimético no son mejores que con el modelo de referencia entonces no es adecuado utilizar la técnica mimética. De forma particular, solamente usamos la métrica de calidad cuando la precisión del modelo mimético es cuando menos 1% mejor que el modelo de referencia, con esta restricción siempre tendremos una métrica de calidad positiva y por lo tanto comparable.

Como veremos, los factores que afectan a la métrica de calidad Q son múltiples y complejos. Mostraremos que usando la poda tradicional en el árbol de decisión no da mucho margen para mejorar esta función de calidad. Afortunadamente, la manera en la cual se genera el conjunto de datos aleatorio puede dar más maniobrabilidad para mejorar la métrica. El uso apropiado de un tamaño del conjunto de datos aleatorios es altamente relevante al tamaño del clasificador mimético. La eliminación de los ejemplos con confianza baja y de la duplicación de ejemplos con alta confianza es también una nueva opción importante que se introduce y se analiza en este artículo. El estudio experimental se ha realizado sobre 20 conjuntos de datos del repositorio UCI [1], y se han considerado dos clases de oráculos: Redes Neuronales y Boosting, usando las implementaciones en el paquete de minería de datos Weka (MultilayerPerceptron

y AdaBoostM1, respectivamente). También, los clasificadores de referencia y mimético son construidos con el algoritmo J48 de Weka. Para todos ellos se ha usado su configuración por defecto. Así mismo, en las gráficas se muestran los resultados de los experimentos obtenidos mediante validación cruzada de diez veces (10-fold cross-validation).

Este documento está organizado como sigue. La sección 2 explica en detalle la técnica mimética. La sección 3 estudia la influencia de la poda en la calidad y la fidelidad de los clasificadores miméticos usando redes neuronales y boosting como oráculos. La relación entre el tamaño del conjunto de datos inventado y la calidad del modelo mimético se analiza en la sección 4. La sección 5 modifica al conjunto de datos inventado tomando en cuenta la confianza del oráculo al etiquetar los ejemplares. La sección 6 combina todos estos factores para buscar el caso óptimo. Finalmente, la sección 7 incluye las conclusiones y el trabajo futuro a realizar sobre el método mimético.

2 El Método Mimético

Domingos presentó en [5][6][7] una aproximación para convertir cualquier modelo en un conjunto de reglas. El método fue bautizado como *CMM* (Combined Multiple Models) y consiste en dos etapas como se muestra en la Figura 1.

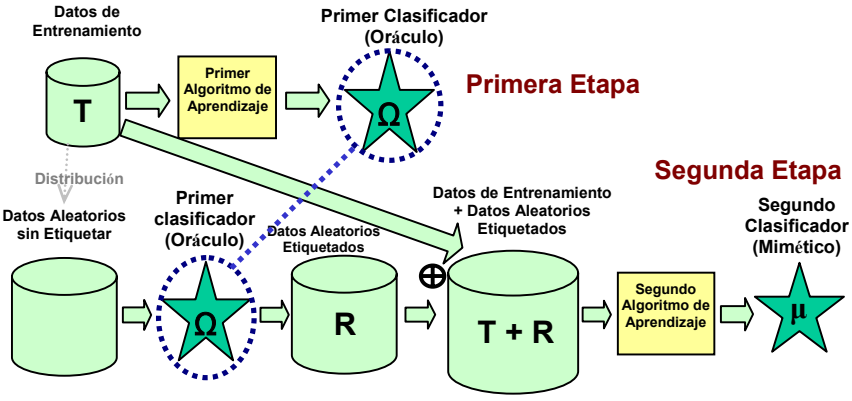

Fig.1. Técnica Mimética

La primera etapa corresponde al aprendizaje del oráculo. El objetivo de esta etapa es obtener un modelo de caja negra. Para ello, se utilizan los datos de entrenamiento originales *T* como entrada al primer algoritmo de aprendizaje, el cual creará este primer clasificador, que es preciso pero incomprensible. Ejemplos de algoritmos que se pueden utilizar en esta primer etapa son: las redes neuronales, boosting, bagging, multiclasificadores (cualquier combinación de clasificadores) y máquinas de vectores soporte.

En la segunda etapa se realiza el aprendizaje mimético. El objetivo de esta etapa es obtener un modelo que imite el comportamiento del modelo de caja negra, buscando las propiedades de conservar el nivel de precisión obtenido con el

modelo de caja negra y generar un conjunto de reglas comprensibles que permitan el entendimiento de dicho modelo. Tomando como base los datos de entrenamiento originales y empleando algún método simple de muestreo se genera un conjunto de datos aleatorio sin etiquetar. Luego ese conjunto de datos *R* se etiqueta utilizando el oráculo. Los conjuntos de datos etiquetados por el oráculo y el conjunto de datos de entrenamiento original se unen para formar el total de los datos de entrada *T+R* al segundo algoritmo de aprendizaje y así crear el modelo mimético. El segundo algoritmo de aprendizaje puede ser cualquier clasificador que genere reglas, como por ejemplo J48 (C4.5[15] en WEKA) y C5. En particular, Domingos empleó bagging [5] como oráculo y C4.5rules como el modelo comprensible final.

Para la generación del conjunto de datos inventado usamos la técnica propuesta en [8]. Básicamente, procedemos como sigue: cada atributo X_i de un nuevo ejemplo se obtiene como el valor v_i en un ejemplo diferente $e(v_1,...,v_i,...,v_m)$ seleccionado del conjunto de entrenamiento por el uso de una distribución a priori. Este procedimiento de generar instancias asume que todos los atributos son independientes. Consecuentemente, el método mantiene las probabilidades de aparición de los diferentes valores observados en cada atributo del conjunto de datos de entrenamiento.

En [9] analizamos experimentalmente el método mimético pero con un oráculo diferente, boosting, obteniéndose unos resultados consistentes con los de Domingos. Adicionalmente, en [9] analizamos teóricamente la técnica, probando que el 100% de fidelidad se alcanza con árboles de decisión sin podar, si se genera un conjunto de datos aleatorios suficientemente grande.

Con todos estos resultados, hemos aprendido algunas cosas acerca de cómo los clasificadores miméticos trabajan:

- Se necesita un gran número de ejemplares aleatorios para llegar a obtener una fidelidad alta, pero a cambio el número de reglas es también alto.
- En [8] se demuestra que se obtienen mejores resultados cuando se usa la distribución a priori para generar el conjunto aleatorio, en comparación con una distribución uniforme.
- El uso del conjunto de datos de entrenamiento original (unido con el conjunto de ejemplares aleatorios) es beneficioso desde el punto de vista de la precisión del modelo mimético.
- Bajo unas pocas suposiciones razonables (el oráculo no es fractal y el mimetizador puede ajustarse al detalle que se quiera), un árbol de decisión no podado (con un conjunto de datos inventado arbitrariamente grande) puede alcanzar el 100% de fidelidad respecto del oráculo.

Resumiendo, el método mimético es un método simple y general para convertir cualquier modelo no comprensible en un conjunto de reglas, con una precisión similar. Sus principales ventajas son:

- Versatilidad: funciona con cualquier "oráculo" (cualquier modelo complejo). Puede implementarse fácilmente en cualquier paquete de minería de datos.
- Trabaja particularmente bien cuando el "oráculo" es una combinación de clasificadores. Es, por tanto, el primer método general para convertir métodos multiclasificadores en modelos simples.

Una desventaja es que como el "oráculo" suele ser complejo, podar excesivamente en la segunda fase anula la mejora en precisión. Por lo tanto, se debe buscar un compromiso entre precisión y tamaño del modelo mimético. Este trabajo se centra en este compromiso

3 Análisis de la Poda

En esta sección analizamos la influencia del uso de diversos grados de la poda en el algoritmo Mim sobre la métrica de calidad y la fidelidad del clasificador mimético. Para hacer esto, realizamos varios experimentos que modifican el factor de poda en el algoritmo J48: desde 0.0001 hasta 0.3. Específicamente, hemos considerado los valores siguientes: 0.0001, 0.001, 0.01, 0.02, 0.05, 0.10, 0.15, 0.20, 0.25, y 0.3. En todos los casos, hemos fijado el tamaño del conjunto inventado de datos aleatorios a 10,000 ya que ésta es una talla lo suficientemente grande para evitar que los resultados se vean influidos por el tamaño de la muestra.

La Figura 2 muestra los resultados medios obtenidos para todos los conjuntos de datos cuando el oráculo es una red neuronal (Mim RN) y cuando el oráculo es boosting (Mim Boost). También hemos incluido como referencia la precisión obtenida por los oráculos y por el algoritmo J48 aprendido con el conjunto de datos de entrenamiento original.

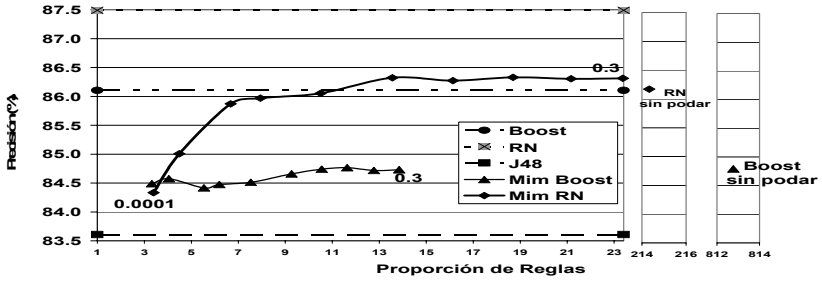

Fig. 2. Precisión contra Proporción de Reglas de Mim para diferentes grados de poda

Como podemos ver, Mim tiene un comportamiento mejor cuando el oráculo es la red neuronal. La razón es que la red neuronal es en promedio un oráculo mejor que el boosting (en términos de precisión). También, ambos son mejores que el J48. Sorprendentemente, Mim Boost presenta un aumento muy pequeño en la precisión con respecto al J48, que se podría explicar por el hecho de que el J48 puede ser visto en sí mismo como una clase de algoritmo boosting [11]. En la Figura 2 podemos ver que el aumento en precisión alcanza prácticamente su máximo con un grado de la poda de 0.01 y con una proporción de reglas alrededor de 7 veces más que el clasificador de referencia J48. Este punto "óptimo" es corroborado por Figura 3, la cual muestra la métrica de calidad para Mim RN.

Si consideramos solamente aquellos conjuntos de datos para los cuales el oráculo mejora la precisión del algoritmo J48, ver Figura 4, el uso de un clasificador mimético en vez de un árbol simple de decisión se justifica mejor. Por ello, analizamos la fidelidad (% de coincidencias entre el oráculo y el modelo mimético) en vez de la precisión. En estos casos el oráculo mejor es boosting.

Esto se debe al hecho de que es más fácil imitar un oráculo como boosting que se base en la misma clase de regiones (eje paralelo) que un oráculo como la red neuronal cuyas regiones son más diferentes. En los casos mejores el rango de proporción de reglas es más corto que el de los peores casos, esto debido a la mayor rapidez de crecimiento de la fidelidad.

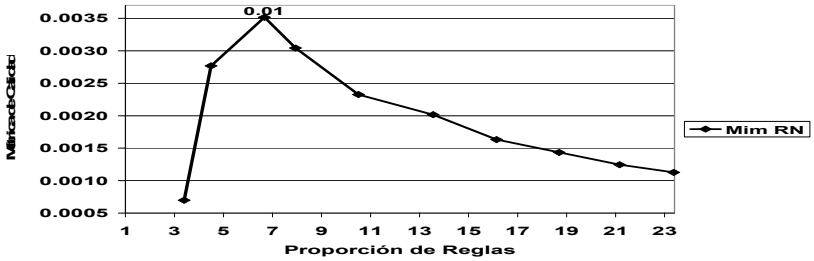

Fig. 3. Quality Metric vs Proportion of rules of Mim RN for several degrees of pruning

De todos los resultados anteriores, se concluye que podando, o por lo menos el método de poda incluido en J48, da una pobre maniobrabilidad para conseguir buenos resultados de la precisión y tener pocas reglas. Las mejores calidades se obtienen con un aumento de casi 2 puntos en precisión pero con una proporción de reglas de 7, es decir los árboles de decisión son 7 veces más grandes que los originales. Esto genera un problema de comprensibilidad, dada la gran cantidad de reglas. Por lo tanto, en la sección siguiente, estudiamos la influencia de otros factores más interesantes tales como el tamaño inventado del conjunto de datos.

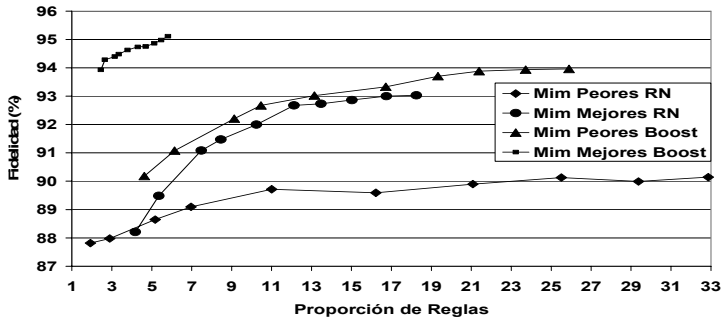

Fig. 4. Fidelidad contra Proporción de Reglas de Mim para varios grados de poda cuando el oráculo excede en precisión al J48

4 Análisis del Tamaño del Conjunto de Datos Inventado

Los trabajos previos sobre los clasificadores miméticos [5][6][7][8] han considerado un tamaño fijo para el conjunto de datos inventados (generalmente entre 1,000 o 10,000). Sin embargo, está claro que este valor es relativamente pequeño para conjuntos de datos tales como "setter" (cuyo conjunto de datos consta de 20,000 ejemplares) y relativamente grande para los conjuntos de datos tales como "hayes-roth" (que tiene 132 ejemplares).

En [9], se demostró teóricamente que el método mimético tiene un rango de utilidad con dos posibles situaciones, que se muestran gráficamente en las Figuras

5 y 6. Estas figuras incluyen tres curvas: la curva de aprendizaje para el primer clasificador (oráculo), la curva de aprendizaje para el segundo clasificador (referencia), ambas obtenidas con diferentes tamaños del conjunto de datos, y adicionalmente, la curva del clasificador mimético, considerando un conjunto de datos aleatorio infinito. Aunque estas gráficas son solamente ilustrativas, sirven para ver cuándo el método es inútil (si la curva mimética está debajo de la curva del segundo clasificador) o en el caso de que sea útil, para qué tamaños de entrenamiento es útil. Las curvas de aprendizaje son obtenidas con la función:

$$\text{Precisión} \quad = \left(a - \frac{1}{c}\right)\left(1 - e^{\alpha n}\right) + \frac{1}{c} \qquad (2)$$

donde a es la máxima precisión, α es la tasa de crecimiento, n tamaño del conjunto de datos de entrenamiento y c es el número de clases (5 en estas figuras).

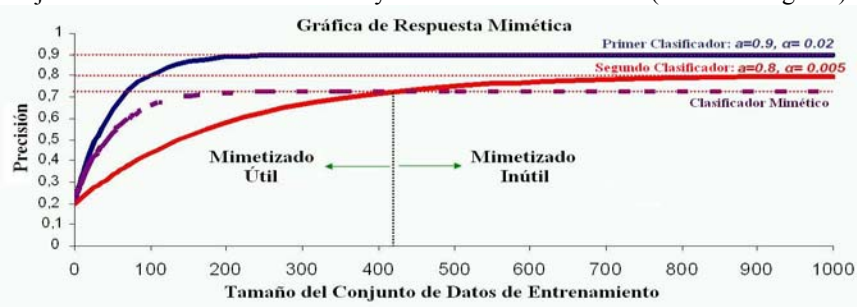

Fig 5. Gráfica de respuesta mimética cuando el método es útil.

La figura 5 muestra un caso donde el método mimético es útil para conjuntos de datos de entrenamiento pequeños (obteniendo de la gráfica el número de ejemplares <=415), porque para valores mayores es preferible usar el clasificador comprensible directamente.

La figura 6 muestra el caso cuando el método mimético es inútil independientemente del tamaño del conjunto de datos de entrenamiento, ya que siempre está por debajo del clasificador de referencia.

Fig. 6 Gráfica de respuesta mimética cuando el método es inútil.

Para estudiar con mayor profundidad la relación entre el tamaño del conjunto de datos inventado y la calidad del modelo mimético en términos de la métrica de calidad, realizamos experimentos con varios tamaños del conjunto de datos

inventado, de 0.3n a 6n, siendo n el tamaño del conjunto de datos de entrenamiento original. Cada incremento fue de 0.3, haciendo un total de 20 tamaños diferentes por cada conjunto de datos.

La Figura 7 muestra la métrica de calidad en función de la proporción de reglas para cada tamaño del conjunto de datos inventado considerado. Como podemos ver, hay un punto máximo en (3.06, 0.0071), con un conjunto de datos inventado 1.5 veces más grande que el conjunto de datos de entrenamiento, lo cual significa que, como esperábamos, conjuntos de datos pequeños tienen una precisión muy baja pero, por otro lado, tampoco es beneficioso generar conjuntos de datos demasiado grandes.

Puesto que la precisión no crece linealmente, está claro que tenemos un punto de saturación para la métrica de calidad como el que se muestra en la Figura 7. Sin embargo, esta saturación no se alcanza en el mismo punto para cada conjunto de datos. Por lo tanto, considerar 1.5 como valor para todos los conjuntos de datos puede no ser siempre una buena opción.

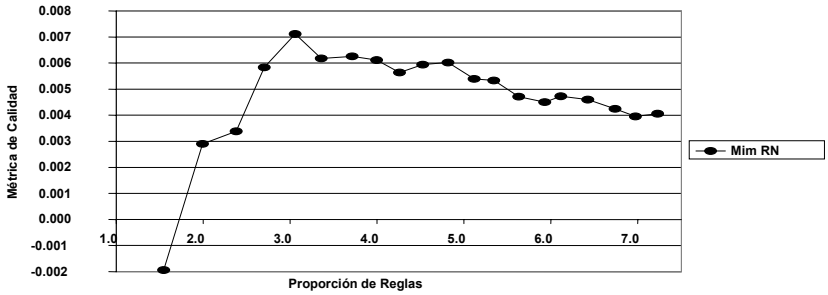

Fig. 7. Métrica de Calidad contra Proporción de Reglas de Mim para diferentes tamaños

En [2] se hizo un estudio experimental para calcular cuál es el tamaño óptimo del conjunto de datos inventado para cada problema. La formula obtenida es

$$n = \varphi \times Tamaño$$

dónde el factor φ (tamaño aleatorio/tamaño entrenamiento) *se* calcula como

$$\varphi = 0.05097 \times AtrNum - 0.01436 \times AtrNom + 0.17077 \times Clases + 0.00003 \times Tamaño$$

siendo *AtrNom* el número de atributos nominales, *AtrNum* el número de atributos numéricos, *Clases* el número de clases y *Tamaño* el tamaño del conjunto de datos de entrenamiento original. Usaremos este valor n en los experimentos realizados en la sección 6.

5. Influencia del Umbral de Confianza y del Grado de Repetición de los Ejemplos

Otra cuestión que analizamos es cómo influye la confianza de los ejemplos en la técnica mimética. La idea es reducir el tamaño del conjunto de entrenamiento usado para generar el modelo mimético considerando únicamente aquellos datos para los que el oráculo tiene una cierta confianza (probabilidad de la clase predicha).

Para hacer esto, procesamos al conjunto de datos $R+T$ (siendo R el conjunto inventado y T el conjunto de entrenamiento original) eliminando los ejemplares

cuyo valor de confianza esté por debajo de un umbral de confianza t_c. Nótese que los ejemplos de T nunca serán eliminados porque tienen un valor de confianza de 1. Después, quitamos los ejemplos repetidos. Finalmente, duplicamos los ejemplos restantes cierto número de veces dependiendo de su valor de confianza. El conjunto de datos que resulta, que lo denotamos como D_{RT}, se utiliza para entrenar al modelo mimético. El número de veces que un ejemplo debe ocurrir en D_{RT} se define como sigue: dados el valor de confianza C_e de un ejemplo e, y un factor F de repetición, entonces el número de ocurrencias de e en D_{RT} es $occ(e)$= $round(C_e \times F)$. La Figura 8 ilustra este proceso

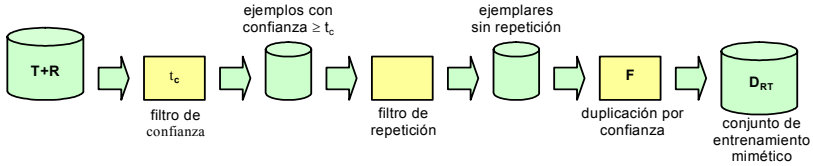

Fig. 8. Trasformación del conjunto de entrenamiento para la técnica mimética

Para los experimentos los umbrales de confianza t_c usados son 0, 0.3, 0.9, 0.95, 0.98, 0.99 y 1.0, y para cada uno de ellos se utiliza un factor de repetición que varía de 1 a 4. El tamaño del conjunto de datos inventado R es de 10,000, que es una muestra suficientemente grande. La Figura 9 muestra estos resultados. La precisión del oráculo (RN) así como la precisión del método mimético original (es decir, sin procesamiento del conjunto de datos inventado, Mim) también se han incluido como referencia.

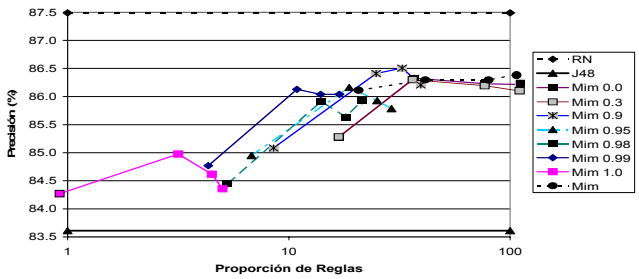

Fig. 9. Precisión contra Proporción de reglas de Mim dependiendo de un umbral de confianza y un factor de repetición

Observamos que en el caso del umbral de confianza t_c =1.0 y de un factor de repetición F=1, en general el proceso no agrega ejemplos inventados por lo que no lo tendremos en cuenta en la discusión.

Para el caso de t_c=0.99 y de F=1, conseguimos un tamaño de ejemplos inventados en D_{RT} alrededor de 3,000. Si contrastamos este valor con un tamaño de ejemplos inventados en D_{RT} alrededor de 7,000 cuando utilizamos t_c=0.0 (se eliminaron aproximadamente 3,000 ejemplos inventados porque se repiten), podemos ver que un porcentaje importante de ejemplos dan una confianza igual o mayor a 0.99 para la RN (cerca del 50%). Consecuentemente, con t_c=0.99 tenemos una situación intermedia la cual está más a la izquierda en la Figura 9 que el Mim original (lo que significa que se obtiene una mejor precisión con una proporción

de reglas menor). Adicionalmente, el aumento en precisión se alcanza casi totalmente con este caso (86.1).

Con respecto al factor de repetición, el comportamiento es absolutamente similar para todos los casos, pero tiene diversas interpretaciones. Por ejemplo, para t_c=0.99 y para F=1 todos los ejemplos restantes son incluidos una vez y para F=2 todos los ejemplos restantes se incluyen dos veces. El aumento importante de la precisión entre estos dos casos se puede justificar por el hecho de que J48 tiene en su configuración por defecto una limitación en el número mínimo de ejemplos por nodo, y esta duplicación permite que J48 sea más preciso.

Para confirmar estas aseveraciones mostramos en la Figura 10 la métrica de calidad para estos experimentos. Como podemos ver, la mejor métrica de calidad se obtiene usando un umbral de confianza de 0.99 (Mim 0.99) con un factor de repetición de 2.

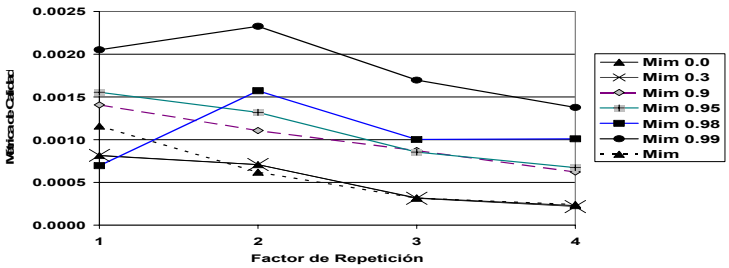

Fig. 10. Métrica de Calidad contra Factor de Repetición para diferentes umbrales de confianza

6. Búsqueda del Caso Óptimo por Combinación de Factores

Finalmente, hicimos un experimento que combinaba algunos de los resultados obtenidos en los experimentos anteriores. Utilizamos un factor de poda de 0.01, el tamaño del conjunto de datos inventado fue fijado al valor predicho según el modelo obtenido en la sección 4, el factor de repetición se fijó a 2 y el nivel de confianza a 0.99. Dado que todos estos factores no son independientes entre sí, también se realizaron experimentos con un factor de poda de 0.1, puesto que la combinación de varios factores para reducir el número de reglas podría obtener un resultado total con un número corto de reglas pero también con una precisión baja. La Tabla 1 muestra los resultados de ambos escenarios.

Los resultados con el nivel de poda de 0.01 muestran que los tres factores principales considerados (poda, tamaño del conjunto de datos y relevancia de los ejemplos), si se utilizan juntos, pueden reducir dramáticamente el número de reglas. De hecho, los resultados medios muestran que el número de reglas del modelo mimético es inferior al número de reglas del modelo de referencia (J48 con su configuración por defecto). En este escenario, sin embargo, el aumento en la precisión es suave (de 83.61 a 84.36). El cuadro cambia cuando vemos los resultados con el nivel de poda de 0.1. En este caso, la precisión aumenta hasta 84.68 con un tamaño de los modelos que es solamente 2.52 veces más grande que el modelo original J48.

Tabla 1. Resultados experimentales obtenidos por la combinación de varios factores

No.	RN	J48		Mim 0.01			Mim 0.1		
Dataset	Acc	Acc	Rules	Acc	Rules	Ratio	Acc	Rules	Ratio
1	98.89	98.56	39.50	98.18	48.60	1.23	98.39	53.05	1.34
2	83.21	77.33	30.20	85.05	53.30	1.77	85.76	53.10	1.76
3	90.84	78.40	39.60	77.88	32.83	0.83	79.60	52.45	1.32
4	67.96	74.08	7.50	70.39	3.17	0.42	72.17	19.15	2.55
5	50.86	51.57	155.70	54.51	47.67	0.31	51.69	228.55	1.47
6	81.94	85.13	5.50	84.40	5.20	0.95	85.44	6.25	1.14
7	74.42	74.19	19.20	74.53	27.20	1.42	72.53	63.50	3.31
8	81.20	68.58	19.00	74.74	22.63	1.19	77.25	24.65	1.30
9	80.06	79.43	9.40	79.35	2.27	0.24	79.63	5.55	0.59
10	96.81	94.96	4.70	95.33	4.77	1.01	94.67	4.65	0.99
11	82.08	87.98	1,158.10	87.65	1,037.20	0.90	85.69	16,655.10	14.38
12	100.00	97.12	30.10	100.00	28.00	0.93	100.00	28.00	0.93
13	100.00	63.29	24.50	65.72	1.00	0.04	65.72	1.00	0.04
14	98.49	98.92	14.00	96.95	9.97	0.71	98.92	13.70	0.98
15	100.00	100.00	25.00	99.90	90.10	3.60	99.98	161.70	6.47
16	96.84	98.68	28.60	98.32	10.65	0.37	98.32	10.65	0.37
17	94.49	96.55	5.80	95.49	2.33	0.40	96.22	5.70	0.98
18	93.15	79.75	128.00	79.50	155.75	1.22	82.93	494.9	3.87
19	95.02	92.39	8.30	92.82	12.03	1.45	92.68	14.15	1.70
20	83.54	75.36	290.70	76.54	177.10	0.61	76.02	1450.2	4.99
Avg.	87.49	83.61		84.36		0.98	84.68		2.52

Estos resultados son perceptiblemente mejores que los obtenidos por los trabajos previos sobre los clasificadores miméticos [5][6][7][8]. Aunque los conjuntos de datos eran diferentes y el oráculo usado era diferente también, tenemos, por ejemplo, que en [5] se obtuvo un aumento de 1.5 puntos en precisión (de 75.9 a 77.4) con una proporción de reglas de 4.9, teniendo una calidad de 0.0051. En la sección 4 con un tamaño de conjunto de datos inventado del 330% se obtuvo un caso similar con una proporción de reglas (4.68), 2.4 puntos de aumento en precisión (de 83.61 a 86.01) y una calidad de 0.006. Y acabamos de ver que podemos obtener aumentos significativos con solamente una proporción de reglas de 2.52 y una calidad de 0.0084.

7. Conclusiones y Trabajo Futuro

Resumiendo, de trabajos previos y después del análisis en algunos de los factores separados (especialmente la poda), parecía que era casi imposible mejorar la métrica de calidad. La reducción del número de reglas implica sistemáticamente una reducción de la precisión y viceversa. Sin embargo, el estudio de factores tales como el tamaño del conjunto de datos inventado y de la modificación de la distribución de ejemplos (por selección y repetición) son herramientas mejores para mantener mejoras significativas en precisión que perceptiblemente reducen el número de reglas.

Con los resultados que se obtuvieron en los primeros trabajos sobre el método mimético, la aplicabilidad de esta aproximación para problemas reales era limitada por el excesivo número de reglas del modelo mimético. Los nuevos análisis y resultados demuestran que la técnica mimética permite extraer conjuntos de reglas cortos a partir de cualquier modelo de caja negra, y usar este conjunto de reglas como la explicación de la caja negra.

Adicionalmente, el trabajo proporciona otro avance en cómo los clasificadores miméticos trabajan. Una de las nuevas contribuciones principales del trabajo es la

modificación de la distribución de los ejemplos, utilizando el nivel de confianza de la clase predicha por el oráculo. Se piensa que el aumento en el número de reglas puede ser parcialmente debido al sobreajuste de los ejemplos de baja confianza etiquetados por el oráculo.

Finalmente, como trabajo futuro, quisiéramos investigar en varias direcciones. Por ejemplo, los métodos de selección de instancias podrían ser útiles para reducir el tamaño del conjunto de datos inventado. La evaluación de modelos miméticos con otras métricas, tales como el AUC (área debajo la curva ROC), también sería interesante, puesto que los árboles de decisión tienen AUCs mejores cuando el árbol no se poda [14]. Otra asepecto a estudiar sería analizar el uso de la confianza sin el conjunto de datos de entrenamiento, buscando hacer a la técnica mimética más generalmente aplicable.

Referencias

1. Black C. L.; Merz C. J.UCI repository of machine learning databases, 1998.
2. Blanco-Vega R., Hernández-Orallo J., Ramírez-Quintana Ma. J. Analysing the Trade-off between Comprehensibility and Accuracy in Mimetic Models. *The 7th International Conference on Discovery Science*, 2004.
3. Boz, O. Converting A Trained Neural Network To A Decision Tree Dectext Decision Tree Extractor, Thesis, Comp. Science and Eng. Dep., Lehigh University (2000)
4. Craven, M. W.; Shavlik, J. W. Using Sampling and Queries to Extract Rules from Trained Neural Networks, *Proc. of the 11th Int. Conf. on Machine Learning*, pp: 37-45, 1994.
5. Domingos, P. Knowledge Discovery Via Multiple Models. *Intelligent Data Analysis*, 2(1-4): 187-202, 1998.
6. Domingos, P. Learning Multiple Models without Sacrificing Comprehen-sibility, *Proc. of the 14th National Conf. on AI*, pp:829, 1997.
7. Domingos, P. Knowledge Acquisition from Examples Via Multiple Models. *Proc. of the 14th Int. Conf. on Machine Learning*, pp:98-106, 1997.
8. Estruch, V.; Ferri, C.; Hernandez-Orallo, J.; Ramirez-Quintana, M.J. Simple Mimetic Classifiers, *Proc. of the Third Int. Conf. on Machine Learning and Data Mining in Pattern Recognition*, LNCS 2734, pp:156-171, 2003.
9. Estruch, V.; Hernández-Orallo, J.: Theoretical Issues of Mimetic Classifiers, TR DSIC, http://www.dsic.upv.es/~flip/papers/mim.ps.gz, 2003.
10. Johnson, G.; Nealon, J. L.; Lindsay, R. O. Using relevance information in the acquisition of rules from a neural network, *Proc. of the Workshop on Rule Extraction from Trained Artificial Neural Networks*, pp: 68-80, 1996.
11. Kearns M.; Mansour Y. On the boosting ability of top-down decision tree learning algorithms. *Proc. of the 28th Annual ACM Symp. on the Theory of Computing*, pp: 459-468, 1996.
12. Núñez, H.; Angulo, C.; Català, A. "Support Vector Machines with Symbolic Interpretation", 7th Brazilian Symp. on Neural Networks, pp:142-149, 2002.
13. Pop, E., Hayward, R., and Diederich, J.: RULENEG: Extracting Rules From a Trained ANN by Stepwise negation, technical report, QUT NRC (1994)
14. Provost, F.; Domingos, P. Tree induction for probability-based rankings, *Machine Learning*, 52 (3): 199-215, 2003.
15. Quinlan, J. Ross. *C4.5: Programs for machine learning*, Morgan Kaufmann Publishers, 1993.
16. Taha, I.; Ghosh, J. Three techniques for extracting rules from feedforward networks, *Intelligent Engineering Systems Through Artificial Neural Networls*, 6: 23-28, 1996.