



Inteligencia Artificial. Revista Iberoamericana
de Inteligencia Artificial

ISSN: 1137-3601

revista@aepia.org

Asociación Española para la Inteligencia
Artificial
España

Barandela, Ricardo; Gasca, Eduardo; Alejo, Roberto

Corrección de la muestra para el aprendizaje del perceptrón multicapa

Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 5, núm. 13, 2001, pp. 2-9

Asociación Española para la Inteligencia Artificial

Valencia, España

Disponible en: <http://www.redalyc.org/articulo.oa?id=92521301>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

CORRECCION DE LA MUESTRA PARA EL APRENDIZAJE DEL PERCEPTRON MULTICAPA.

Ricardo Barandela, Eduardo Gasca, Roberto Alejo

Instituto Tecnológico de Toluca
Ave. Tecnológico s/n , 52140 Metepec, México
rbarandela@hotmail.com

Resumen

Es muy popular, en la actualidad, el empleo de las redes neuronales artificiales, en particular el Perceptron Multicapa (PM), para tareas de reconocimiento de patrones, minería de datos y aprendizaje automático. No obstante, aún se conoce poco de estos modelos, lo que se traduce en debilidades tales como lentitud en el proceso de aprendizaje y pobre capacidad de generalización.

En el presente trabajo se muestra (con datos artificiales y reales) la conveniencia de procesar la muestra de entrenamiento con técnicas tomadas del contexto de la regla NN (Nearest Neighbor rule), para acelerar el entrenamiento e incrementar la precisión del PM. Se ejemplifica, además, como el rendimiento del PM se afecta de manera notable en situaciones imperfectamente supervisadas (o sea, cuando se tienen errores de identificación en algunos patrones de entrenamiento) que suelen ocurrir en la práctica, y como esta deficiencia se puede eliminar si se emplea la metodología de Depuración, que ya ha evidenciado sus beneficios cuando se clasifica con la regla NN.

Palabras clave: Aprendizaje supervisado, Perceptron Multicapa, Regla NN, Patrones Atípicos, Identificaciones incorrectas.

1. Introducción

En los últimos años se ha popularizado el empleo de las redes neuronales artificiales (Bishop, 1996; Ripley, 1996) para tareas de aprendizaje automático, reconocimiento de patrones y de minería de datos. En particular, el modelo conocido como Perceptron Generalizado o Multicapa (Backpropagation) ha sido utilizado en la interpretación de imágenes de percepción remota (p. ej., Wilkinson et al., 1995; Foody, 1995, 2000) y en otras labores de clasificación. No obstante, aún se conoce poco de estos modelos, lo que se traduce en debilidades tales como la lentitud en el aprendizaje y la pobre capacidad de generalización que se observa en un

número importante de aplicaciones prácticas. Estos dos problemas: incrementar la rapidez del procedimiento de entrenamiento del clasificador y mejorar la precisión en el resultado de sus decisiones, han motivado un esfuerzo importante en la investigación de criterios más adecuados para definir parámetros y algoritmos vinculados al proceso de aprendizaje (p. ej., Vitela y Reifman, 1997; Magoulas et al., 1997; Gorse et al., 1997; Abdel-Wahhab y Sid-Ahmed, 1997; Lee, 1997; Foody y Arora, 1997; Atiya y Ji, 1997). Ha resurgido, también, el interés por examinar la calidad de la muestra de entrenamiento (ME) y la validez de sus elementos. En su estudio de las dificultades encontradas en aplicaciones del Perceptron Multicapa (PM) para clasificar datos de

teledetección, (Paola y Schowengerdt, 1995) mencionan que la mayor parte de la convergencia al error mínimo ocurre en las primeras iteraciones del entrenamiento y la proporción de la mejoría en ese error decrece dramáticamente a medida que avanza el aprendizaje. Esos autores citan el trabajo de Rau y Lure (1993), en el que se reportó que una red neuronal clasificó correctamente 83 porciento de la ME en 20 iteraciones, 93.3 porciento en 60 iteraciones y requirió 2000 iteraciones para llegar a identificar 95.5% de los patrones de entrenamiento. Estos hechos parecen apuntar hacia el efecto negativo que la presencia de valores atípicos (outliers) dentro de la ME pueden provocar en la eficiencia y en el rendimiento del clasificador.

En el presente trabajo se exploran posibilidades para modificar y reducir el tamaño de la muestra para el entrenamiento del PM, con el objetivo de eliminar patrones atípicos y corregir posibles identificaciones erróneas de esos patrones de entrenamiento (situaciones imperfectamente supervisadas). Estas propuestas, con las que se pretende combatir la carga computacional del sistema y mejorar su capacidad de generalización, han sido tomadas de la experiencia con otro clasificador no paramétrico, la regla NN. En la sección siguiente se exponen los rasgos básicos de la regla NN y sus analogías con el PM, y se describen brevemente las técnicas propuestas. En una sección posterior se presentan y comentan resultados experimentales con datos artificiales y reales. El trabajo termina con algunas observaciones sobre las líneas de investigación que se proseguirán en esta dirección y se comentan otras propuestas relacionadas con estos temas.

2. La regla NN y técnicas para el procesamiento de la ME

La regla NN (Nearest Neighbor rule o regla del vecino más cercano), de gran popularidad entre teóricos y especialistas de los campos de aplicación, es un clasificador que comparte con el Perceptrón Multicapa varias características:

- es un método supervisado, lo que implica la disponibilidad (y el empleo) de una muestra de entrenamiento, que se supone compuesta por patrones perfectamente identificados y que representan a todas las clases de interés en el problema que se desea resolver.
- es un método no paramétrico, es decir, independiente de todo tipo de modelo probabilístico para los datos.

- adolece de una considerable carga computacional. En el caso de la regla NN, esto está dado por la necesidad, según la definición de la regla, de conservar la ME completa en memoria y de examinar a cada uno de los patrones de entrenamiento para poder clasificar a un patrón desconocido.

Numerosos procedimientos se han desarrollado para disminuir esos requerimientos, en capacidad de almacenamiento y en tiempo de cálculo, a niveles aceptables. Estos esfuerzos han seguido dos direcciones distintas, una de las cuales consiste en la reducción del tamaño de la ME mediante la eliminación de algunos de sus elementos, pero conservando el máximo posible de la información discriminatoria. La idea de Hart (1968) sobre el subconjunto consistente ha servido de inspiración a numerosos investigadores. El subconjunto consistente se define como aquel subconjunto (de la ME) que garantiza 100% de clasificación correcta con la regla NN de todos los patrones de entrenamiento. Sin embargo, el algoritmo propuesto por Hart adolece de varias deficiencias: dependencia del orden en que se evalúan los patrones de entrenamiento, eliminación de algunos de los puntos cercanos a las fronteras de decisión entre las clases y conservación innecesaria de algunos patrones superfluos. Wilson y Martinez (2000) revisan un número considerable de propuestas publicadas con el objetivo de enmendar esas deficiencias. En el presente trabajo se utilizó el método Selectivo Modificado (Barandela et al., 2000) por ser el que logra una mejor aproximación a las fronteras de decisión definidas por la ME completa, lo que se traduce en una superioridad en el nivel de precisión de clasificación de la regla NN tal como se ha demostrado en comparaciones experimentales con otros 13 algoritmos utilizando datos reales.

Aunque con frecuencia se reporta como otra técnica para la reducción del tamaño de la ME, por tener también esa propiedad, el verdadero objetivo de la Edición (Wilson, 1972) consiste en incrementar el porcentaje de clasificación correcta. Para lograr esto, el procedimiento elimina aquellos patrones de entrenamiento (para cada uno de ellos se buscan sus tres vecinos más cercanos en el resto de la ME y se sacan aquellos cuya etiqueta no coincide con la de la mayoría de esos tres vecinos) que se ubican en la zona de solape entre las clases y que, por esa razón, pueden ser considerados como patrones atípicos. También se han publicado diversas modificaciones o variantes de la Edición y todas ellas han mostrado una disminución importante de los errores de clasificación, pero la reducción alcanzada en el tamaño de la ME no ha sido apreciable: sólo se

elimina un 20-25% de los patrones de entrenamiento.

Resultados sorprendentes se han obtenido cuando se combina la Edición con el Selectivo Modificado, aplicados en ese orden, pues se han registrado reducciones de más de 90% en el tamaño de la ME y, al mismo tiempo, mejorías en la precisión del clasificador en comparación con la lograda al trabajar con la ME completa (Barandela, 1987). La razón para este comportamiento de la combinación Edición-Reducción ha sido explicada (Devijver y Kittler, 1982; Sánchez et al., 1997) aduciendo que la Edición deja la ME en las condiciones adecuadas para satisfacer los supuestos implícitos de los algoritmos de reducción basados en el concepto de subconjunto consistente: no solape entre las clases y no presencia en la ME de patrones ruidosos (atípicos). Estrategias como ésta o similares han permitido el empleo de la regla NN en tareas con grandes volúmenes de datos, como ocurre en la interpretación de imágenes de teledetección (Hardin, 1994; Cortijo y Pérez de la Blanca, 1996; Barandela y Castellanos, 1996).

Como ya se mencionó, los métodos supervisados descansan en el supuesto de que todos los elementos de la muestra de entrenamiento están perfectamente identificados (por un experto humano) como miembro de una de las clases consideradas. Este supuesto no se cumple siempre en la práctica pues la recolección de una ME resulta, en muchas aplicaciones, costosa y de notable dificultad. De aquí que con frecuencia se produzcan situaciones conocidas como imperfectamente supervisadas, en las que no es posible garantizar que es correcta la identificación de todos los patrones de entrenamiento. Se ha comprobado que estas situaciones afectan de manera considerable la precisión del clasificador. Barandela y Gasca (2000) presentan una metodología para la restructuración de la ME que provoca cambios en la identificación (traslados de una clase a otra) en algunos de sus miembros y eliminación de otra parte de ellos. Esta metodología (Depuración de la ME) que consiste en la aplicación reiterada de la Edición Generalizada (Koplowitz y Brown, 1978), posteriormente el empleo de la Edición –quizás también más de una vez- y al final el Selectivo Modificado cuando el tamaño de la ME lo requiere, ha sido utilizada con éxito (Barandela, 1995) también en situaciones no supervisadas (como un complemento del algoritmo de agrupamiento o cluster, para refinar los resultados de éste) y en aplicaciones de las geociencias (campo en el que suele ser difícil definir con exactitud la separación entre clases, tanto física como conceptualmente) supuestamente bien supervisadas.

3. Resultados Experimentales con el PM.

Para evaluar las posibilidades de la estrategia edición-reducción en la aceleración del proceso de aprendizaje del Perceptron Multicapa, se desarrollaron experimentos con datos simulados según un modelo muy utilizado en la literatura especializada: dos clases con distribuciones gaussianas bivariadas y con distancia de Mahalanobis entre las clases igual a 2.58. Se generaron pseudoaleatoriamente 10 juegos de datos, cada uno con 100 patrones de entrenamiento para cada clase, y una muestra independiente o de control (MC) con 250 patrones de cada población. A cada una de esas MEs se les aplicó, primero la Edición y después el algoritmo Selectivo Modificado, dando lugar a 30 juegos de datos (ME original:MO, ME editada: MED, y ME editada-reducida:MER, con 10 réplicas en cada caso) para entrenar la red neuronal que se utilizó posteriormente para clasificar la MC. La red se programó en lenguaje C (una versión simplificada del Anexo A en el texto de Pao, 1989) con una capa interna (hidden). Se definieron los parámetros: Razón de aprendizaje = 1.0 y Factor Momentum = 0.8. El criterio de parada para el entrenamiento se estableció como un error general menor a 0.01, o un número máximo de iteraciones igual a 20000. Los resultados obtenidos, en valor promedio para las 10 repeticiones de cada tipo, se muestran en la Tabla No. 1.

| ME utilizada | Tamaño | No.iteraci. | % error |
|---------------|--------|-------------|---------|
| Original (MO) | 200 | 20000 | 19.4 |
| Editada (MED) | 171 | 5 | 24.6 |
| E-Red. (MER) | 11 | 51 | 9.1 |

Tabla No. 1. Experimentos del PM con datos simulados. Valores promedio de las 10 réplicas

En ninguna de las 10 repeticiones con la MO se logró convergencia de los pesos de la red antes de los 20000 iteraciones (con algunos de esos juegos de datos se probó hasta 30000 con el mismo resultado). Por otra parte, en 4 de las 10 repeticiones con MER se llegó también hasta el límite del total de iteraciones pero, dada la reducción en el tamaño de la ME, esas 20000 iteraciones se lograban en menos de un minuto en lugar de los 18-19 que requería la MO. En las cifras de promedio de iteraciones y de precisión de las MER no se tomaron en cuenta esos 4 casos, pero sí fueron considerados al calcular los porcentajes de error de clasificación, con la misma muestra de entrenamiento y con la MC.

Resulta evidente la mejoría conseguida al aplicar edición-reducción a la ME antes de emplear ésta para ajustar los pesos del Perceptrón Generalizado. Y esta mejoría se refleja no sólo en un entrenamiento más rápido sino también en una mejor precisión y capacidad de generalización de la red.

Para estudiar los efectos de la contaminación en la ME (imperfectamente supervisada) se retomaron los datos artificiales. Para simular una situación imperfectamente supervisada se seleccionaron aleatoriamente 20 patrones (20%) que en realidad representaban a la clase 1 y se identificaron como miembros de la clase 2 en cada una de las MEs utilizadas. Se les aplicó entonces la metodología de Depuración: Edición Generalizada, que con estos datos fue necesario reiterar -en promedio- 6 veces, Edición y por último el algoritmo Selectivo Modificado, dando lugar a 40 juegos de datos para entrenar la red neuronal, que se utilizaba en cada caso para clasificar la MC. Los resultados, en valores promedio de las 10 réplicas ejecutadas, se muestran en la Tabla No. 2. El 70% de los patrones que con toda intención habían recibido una etiqueta inicial incorrecta, fueron devueltos a su clase verdadera por la Edición Generalizada reiterada, mientras que esta técnica no produjo eliminación de ningún prototipo. Algunos de esos patrones de entrenamiento intencionalmente mal identificados y cuyas etiquetas no fueron enmendadas por la Edición Generalizada, fueron eliminados de la ME por la Edición o por el Selectivo Modificado.

| ME utilizada | Tamaño | Iteraciones | % error |
|--------------------|--------|-------------------|---------|
| Orig. Contaminada | 200 | 20000 | 30.0 |
| Edic. Generalizada | 200 | 10 ^a | 23.4 |
| Edición | 187 | 629 ^b | 20.3 |
| Selectivo Modific. | 16 | 2195 ^c | 11.2 |

^a corresponde a 4 réplicas, en las otras 6 no se convergió

^b corresponde a 8 réplicas, en las otras 2 no se convergió

^c corresponde a 5 réplicas, en las otras 5 no se convergió

Tabla No. 2 Perceptron Multicapa en situaciones imperfectamente supervisadas. Datos artificiales.

Los efectos del procesamiento de la ME contaminada, en el porcentaje de error de clasificación con la muestra independiente, son notables (se reduce a poco más de la tercera parte que con la ME original). Es de destacar que la metodología de Depuración conduce a niveles de clasificación erróneas menores que cuando se trabaja con la ME original. En otras palabras, la

descontaminación realizada por la metodología produce niveles de clasificación correcta mayores que cuando se revisa la ME por el supervisor humano, lo que en parte es atribuible a los efectos beneficiosos de la Edición. También apreciable es la disminución en la duración del procedimiento de aprendizaje, tanto por el número de iteraciones como el tiempo invertido en cada una de ellas, dado que la ME decreció, en promedio, 92% en su tamaño.

Los efectos de la metodología de la Depuración sobre el aprendizaje y la capacidad de generalización del PM se comprobaron también en experimentos con datos reales, aunque en este caso no se contó con información acerca del nivel de contaminación presente en los datos ni acerca de si se trataba de situaciones imperfectamente supervisadas o no. Se tomaron siete archivos del Depósito de la Universidad de California at Irvine (Merz y Murphy, 1996). Cada conjunto de datos se particionó aleatoriamente en una ME con 40% de los patrones, aproximadamente, y una Muestra de Control con el resto de los objetos de entrenamiento. La Tabla 2 resume las características principales de estos conjuntos de datos reales.

| Datos | clases | atrib. | patr. ME | patr. MC |
|---------------|--------|--------|----------|----------|
| Breast-cancer | 2 | 9 | 273 | 410 |
| Glass | 6 | 9 | 86 | 128 |
| Iris | 3 | 4 | 60 | 90 |
| Liver | 2 | 6 | 138 | 207 |
| Sonar | 2 | 60 | 83 | 125 |
| Vehicle | 4 | 18 | 339 | 510 |
| Wine | 3 | 13 | 71 | 107 |

Tabla 3. Características de los datos reales.

Los resultados de estos experimentos se muestran en la Tabla 4. Aquí se presenta, tanto el porcentaje de clasificación errónea con la ME original, como los resultados después de aplicar la metodología de Depuración. Para fines de comparación, la última columna muestra los resultados que se obtuvieron cuando se utilizó solamente Edición -en algunos casos en forma repetida-, es decir, sin el empleo de la Edición Generalizada.

En cinco de los archivos y en promedio, la Depuración logró mejorías en la precisión del clasificador en comparación con la obtenida por la ME original y en el archivo Iris no hubo cambios. Como se conoce, este archivo consta de 3 clases y existe solape entre patrones de entrenamiento de las

clases 2 y 3. Merece la pena mencionar que Depuración cambió la etiqueta de uno de esos patrones mientras que la Edición lo eliminó. También la Depuración logra rendimientos iguales o mejores que los de Edición en cuatro de los archivos. Cabe reiterar que se trata de conjuntos de datos reales, sobre los cuales no se conoce el grado de contaminación que los afecta. En ninguno de los casos se presentó sobre-entrenamiento de la red. El procedimiento de aprendizaje siempre se detuvo al llegar a las 100000 iteraciones, sin que se alcanzara convergencia al error mínimo planteado (0.0001).

| Datos | ME original | Metodología Depuración | Edición |
|---------------|-------------|------------------------|---------|
| Breast cancer | 3.9 | 1.7 | 2.0 |
| Glass | 53.1 | 39.8 | 39.8 |
| Iris | 2.2 | 2.2 | 3.3 |
| Liver | 37.2 | 30.9 | 30.4 |
| Sonar | 64.8 | 59.2 | 67.2 |
| Vehicle | 25.1 | 44.4 | 36.1 |
| Wine | 6.5 | 3.7 | 0.9 |
| Promedios | 27.5 | 26.0 | 25.7 |

Tabla 4. Experimentos con datos reales. Porcientos de clasificación errónea.

4. Discusión y trabajos relacionados

En el presente trabajo se han presentado resultados importantes que validan, con datos artificiales y reales, la utilidad de la estrategia edición-reducción para procesar la ME, no tan sólo cuando se emplea la regla NN, sino también con el Perceptrón Multicapa. Se ha mostrado que se produce una merma notable en la carga computacional de este método y –lo que es muy significativo- se mejora de manera sustancial la precisión del clasificador. Se demuestra, además, lo beneficiosa que resulta la metodología establecida con anterioridad en relación con la regla NN (Barandela y Gasca, 2000) para trabajar con muestras imperfectamente supervisadas, produciendo una limpieza adecuada de la ME y contribuyendo al rendimiento del método de clasificación. Debe observarse al respecto que los errores de clasificación con la ME depurada son menores, incluso que con esa misma ME antes de que se le contaminara para la evaluación. Se confirma lo apuntado por otros autores en el sentido de eliminar elementos atípicos de la ME para acelerar el aprendizaje en el PM y se evidencia que esa meta se puede cumplimentar

logrando, al mismo tiempo, mayor capacidad de generalización.

En los últimos años se ha propuesto reiteradamente el empleo de modelos de las redes neuronales artificiales para reducir la carga computacional de la regla NN (por ej., Decaestecker, 1997; Huang et al., 1995). En el presente trabajo se demuestra la utilidad de trabajar en la dirección contraria, utilizando técnicas basadas en la regla NN para mejorar la eficiencia y la precisión del PM, tanto en casos supervisados como imperfectamente supervisados. Resultaría de interés explorar una tercera posibilidad: procesar la ME con las técnicas inspiradas en las ideas de las redes neuronales (Learning Vector Quantizer, Decision Surface Mapping) para incrementar la eficiencia y el rendimiento del Perceptron Multicapa.

En publicaciones recientes –no tan sólo del aprendizaje automático, sino también del reconocimiento de patrones y minería de datos- se ha observado un renovado y encomiable interés por un procesamiento adecuado de la muestra de entrenamiento. Durante mucho tiempo ha prevalecido el criterio de que la tarea básica en ese procesamiento consiste en la selección y/o transformación de las variables o atributos (reducción de la dimensión) y que eso es suficiente para obtener un modelo supervisado óptimo. No obstante, en la actualidad está tomando auge la opinión de que es necesario prestar mucha más atención a la calidad de la muestra de entrenamiento, entendiéndose por esto la preocupación por la representatividad de sus elementos y por el grado de confianza en sus etiquetas. Esto reviste mayor importancia cuando se trabaja con métodos no paramétricos como los dos discutidos en el presente trabajo. Ocurre con estos métodos que, precisamente por no depender de supuestos probabilísticos, se apoyan solamente en la información suministrada por la ME y son en extremo sensibles a cualquier deficiencia en la calidad y confiabilidad de ésta.

En su tesis (John, 1997), hace mucho hincapié en la limpieza de los datos (de la ME) como un paso muy importante en el proceso de minería de datos y propone la eliminación de patrones de entrenamiento sospechosos de atipicidad, que son definidos como aquellos que no se ajustan al modelo construido por la mayoría de la ME. El procedimiento para esto consiste en la construcción de un árbol de decisión (John trabaja con datos categóricos), ajustándolo a toda la ME, y la posterior poda (pruning) de las ramas que contribuyen de manera no significativa al modelo. Los patrones de entrenamiento asociados con las

ramas podadas son eliminados de la ME y el árbol se construye de nuevo desde el principio, a partir de la ME “limpia”. Con este proceso se busca la simplificación del modelo y el incremento de su precisión en la clasificación.

Por su parte (Guha et al., 1998) afirman que “todo conjunto de datos contiene casi siempre valores atípicos” y definen un procedimiento para detectarlos con un algoritmo no supervisado (clustering) jerárquico. Una línea similar defienden (Gopalakrishnan et al., 1995) en su propósito de eliminar de la ME los patrones que motivan lentitud en el entrenamiento del PM. Sin embargo, estos últimos autores emplean un método de partición sin que aclaren como determinar de antemano el número de grupos a formar, lo que le da un alto grado de subjetividad a su procedimiento, que ilustran con ejemplos un tanto rebuscados. Tampoco muestran preocupación por la capacidad de generalización de la red neuronal que resulta.

En la literatura especializada han aparecido también algunas referencias a problemas imperfectamente supervisados. Aunque en el Cap. 8 de su texto (Bishop, 1996), define el preprocesamiento de la ME como “la transformación de los datos a una nueva representación antes de entrenar una red neuronal” y dedica la mayor parte de la discusión a la reducción de la dimensión, menciona la posibilidad de etiquetas de clase incorrectas, pero sin pasar más allá de esa alusión. En el contexto de una aplicación a la clasificación de datos de percepción remota (Brodley y Friedl, 1999) proponen el empleo de un conjunto de clasificadores para filtrar los patrones de entrenamiento, eliminando de la ME aquellos que el filtro no clasifica correctamente. Afirmando que se basan en la técnica de eliminar atípicos en el análisis de regresión, esos autores buscan una mejoría en la calidad de la ME, mediante la eliminación de patrones mal identificados, con el propósito de incrementar la precisión del clasificador.

Conviene destacar algunos aspectos del trabajo de Brodley y Friedl. Aunque trabajan con datos reales de percepción remota y afirman que los errores en las etiquetas de los patrones de entrenamiento es algo común en esas aplicaciones, para evaluar los beneficios de su propuesta introducen contaminación artificial en su muestra, con diferentes niveles de ruido. El procedimiento que proponen se limita a eliminar patrones y no incluye la posibilidad de cambiar la etiqueta de algunos de ellos (esto último es algo que esos autores se proponen como tarea futura). Los experimentos que reportan demuestran que el procedimiento no alcanza a limpiar los datos de manera adecuada,

pues sólo con un nivel bajo de contaminación (5%) la ME “limpia” obtiene resultados de precisión comparables con los de la ME original (sin contaminar).

Todo esto contrasta con el método de Depuración que se emplea en el presente trabajo, que considera tanto eliminación como cambio de etiquetas en patrones de entrenamiento y que, tal como se muestra aquí, si alcanza a combatir la contaminación en un grado satisfactorio, pues el comportamiento con la ME depurada resulta mejor que con la ME sin contaminar con el PM (al igual que con la regla NN con la cual esta mejoría se ha logrado hasta con 45% de ruido, véase Barandela y Gasca, 2000). En el presente trabajo se introduce contaminación artificial en los datos simulados pues, por la forma en que se generaron esos datos, hay total control en cuanto al nivel de ruido presente en cada momento. Pero cuando se ha trabajado con datos reales, la metodología de Depuración se ha utilizado con la ME tal como ésta ha sido preparada por especialistas del campo de aplicación.

Se aprecia en la literatura la falta de un concepto riguroso y unificado para el término atípico (outliers) que se emplea indistintamente para referirse a:

- datos “ruidosos” con errores originados en el proceso de la medición o del registro.
- patrones nuevos no identificados que surgen en la etapa de clasificación y que no pertenecen a ninguna de las clases representadas en la ME (Muzzolini et al., 1998; Tax y Duin, 1998).
- elementos de la ME incorrectamente identificados, situación que constituye uno de los focos del presente trabajo.

La metodología de Depuración que aquí se discute elimina tanto patrones de entrenamiento con errores en los valores de algunos atributos (ruidosos) como aquellos que son simplemente excepciones de la regla general. Aunque se puede discutir si conservar patrones excepcionales es importante para propósitos de modelación, es evidente que la distinción entre ruidosos y excepcionales no sería fácil sin la intervención de un especialista humano. Además, cuando el interés está en el incremento del poder de generalización y en la simplificación del procedimiento de aprendizaje, toda la evidencia apunta a la conveniencia de eliminar estas excepciones de la ME. Es importante tener en cuenta que si esos patrones son excepciones es porque aparecen con frecuencia relativamente baja.

Por esta razón, el número de errores de clasificación que estas excepciones provocan si permanecen en la ME es, en general, mucho mayor que la cantidad de clasificaciones erróneas que su eliminación puede producir.

Es indudable que se requiere profundizar y ampliar en los estudios mencionados, no tan sólo por la importancia del tema sino también por su vínculo con otras tareas del aprendizaje automático y del reconocimiento de patrones, como los procedimientos para trabajar con una ME incompleta (cuando alguna de las clases existentes no ha sido representada con patrones de entrenamiento). Otro aspecto cuyo estudio está previsto para una etapa inmediata es la investigación acerca del efecto que las probabilidades a priori de las diferentes clases en las muestras de entrenamiento puede ejercer sobre los métodos aquí discutidos.

Referencias

- Abdel-Wahba, O. y M.A. Sid-Ahmed. A new scheme for training feed-forward neural networks, *Pattern Recognition*, 30, 3, 519-524, 1997.
- Atiya, A. y C. Ji. How initial conditions affect generalization performance in large networks, *IEEE Trans. on Neural Networks*, 8, 2, 448-451, 1997.
- Barandela, R. The NN rule: an empirical study of its methodological aspects, *Tesis Doctoral*, Berlin, 1987.
- Barandela, R. Una metodología para el reconocimiento de patrones en la solución de tareas geólogo-geofísicas, *Geofísica Internacional*, 34,4, 399-405, 1995
- Barandela, R. Problemas del aprendizaje para las redes neuronales artificiales, *Memorias del III Taller Iberoamericano de Reconocimiento de Patrones*, La Habana, 1997.
- Barandela, R. y E. Castellanos. La regla NN para la interpretación de imágenes de percepción remota, *Memorias del Tercer Taller Informática y Geociencias*, La Habana, 1996.
- Barandela, R. y E. Gasca. Decontamination of training samples for supervised pattern recognition methods. En: *Advances in Pattern Recognition*, F. Ferri et al. (eds.), *Lecture Notes in Computer Science*, 1876, Springer, 2000.
- Barandela, R., N. Cortés y A. Palacios. The Nearest Neighbor rule and the reduction of the training sample size. Presentado a SNRFAI2001, Castellón. 2000.
- Bishop, C.M. *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1996.
- Brodley, C.E. y M.A. Friedl. Identifying mislabeled training data. *Journal of Artificial Intelligence Research*, 11, 131-167, 1999.
- Cortijo, F.J. y N. Pérez de la Blanca. Image classification using nonparametric classifiers and contextual information, *Int. Arch. of Phot. and Remote Sensing*, XXXI, B3, 120-124, 1996.
- Decaestecker, C. Finding prototypes for Nearest Neighbor classification by means of gradient descent and deterministic annealing, *Pattern Recognition*, 30, 2, 281-288, 1997.
- Devijver, P.A. y J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall, Englewood Cliffs, N.J., 1982.
- Foody, G.M. Using prior knowledge in artificial neural networks with a minimal training set, *Int. Journal of Remote Sensing*, 16, 2, 301-312, 1995.
- Foody, G.M. Image classification with a Neural Network: From completely crisp to Fully-Fuzzy situations. En: *Advances in Remote Sensing and GIS Análisis*, P.M. Atkinson y N.J. Tate (eds), Wiley, 2000.
- Foody, G.M. y M.K. Arora. An evaluation of some factors affecting the accuracy of classification by an artificial neural network, *Int. Journal of Remote Sensing*, 18, 4, 799-810, 1997.
- Gopalakrishnan, M., V. Sridhar y H. Krishnamurthy. Some applications of clustering in the design of neural networks, *Pattern Recognition Letters*, 16, 59-65, 1995.
- Gorse, D., A.J. Shepherd y J.G. Taylor. The new ERA in supervised learning, *Neural Networks*, 10, 2, 343-352, 1997.
- Guha, S., R. Rastogi y K. Shim. CURE: An efficient clustering algorithm for large databases. *ACM SIGMOD International Conference on Management of Data*, Seattle, Washington, 1998.
- Hardin, P.J. Parametric and Nearest Neighbor methods for hybrid classification: a comparison of

- pixel assignment accuracy, Phot. Eng. & Rem. Sensing, 60, 12, 1439-1448, 1994.
- Hart, P.E. The condensed nearest neighbor rule, IEEE Trans on Inf. Theory, IT-14, 515-516, 1968.
- Huang, Y.S., K. Liu y C.Y. Suen. A new method of optimizing prototypes for nearest neighbor classifiers using a multilayer network, Pattern Recognition Letters, 16, 77-82, 1995.
- John, G.H. Enhancements to the Data Mining Process, Ph. D. Thesis, Stanford University, 1997.
- Koplowitz, J. y T.A. Brown. On the relation of performance to editing in Nearest Neighbor rules. En: Proc. 4th Int. Joint Conf. on Pattern Recognition, Japón, 1978.
- Lee, C.W. Training feedforward Neural Networks: An algorithm giving improved generalization, Neural Networks, 10, 1, 61-68, 1997.
- Magoulas, G.D., M.N. Vrahatis y G.S. Androutsakis. Effective Backpropagation training with variable stepsize, Neural Networks, 10, 1, 69-82, 1997.
- Merz, C.J. y P. Murphy. UCI Repository of Machine Learning Databases, University of California Irvine, Department of Information and Computer Science. <http://www.ics.uci.edu/~mlearn/MLRepository.html>. 1996.
- Muzzolini, R., Y.H. Yang y R. Pierson. Classifier design with incomplete knowledge. Pattern Recognition, 31, 4, 345-369, 1998.
- Pao, Y.H. Adaptive Pattern Recognition and Neural Networks, Addison-Wesley, Reading, MA, 1989.
- Paola, J.D. y R.A. Schowendert. A review and analysis of backpropagation neural networks for classification of remotely sensed imagery. Int. J. of Rem. Sensing, 16, 16, 3033-3058, 1995.
- Rau, Y.C. y M.F. Lure. Classification of remote sensing data using partially trained neural networks. En: Proc. IGARSS, Japón, 728-730, Agosto 1993.
- Ripley, B.D. Pattern Recognition and Neural Networks, University Press, Cambridge, 1996.
- Sánchez, J.S., F. Pla y F.J. Ferri. Prototype selection for the nearest neighbour rule through proximity graphs. Pattern Recognition Letters, 18, 507-513, 1997.
- Tax, D.M. y R.P. Duin. Outlier detection using classifier instability. En: Advances in Pattern Recognition, A. Amin et al. (eds.), Lecture Notes in Computer Science, 1451, Springer, 1998.
- Vitela, J.E. y J. Reifman. Premature saturation in Backpropagation Networks, Neural Networks, 10, 4, 721-735, 1997.
- Wilkinson, G.G., F. Ferien e I. Kanellopoulos. Integration of Neural and Statistical approaches in Spatial Data classification, Geographical Systems, 2, 1-20, 1995.
- Wilson, D.L. Asymptotic properties of nearest neighbor rules using edited data sets, IEEE Trans. Syst., Man and Cybernetics, SMC-2, 408-421, 1972.
- Wilson, D.R. y T.R. Martínez. Reduction techniques for Instance-based learning algorithms. Machine Learning, 38, 3, 257-286, 2000.