

J. M. Jurado¹, A. Rodero¹, J. M. Caridad², A. Molina¹

¹ Departamento de Genética, Universidad de Córdoba. Campus de Rabanales. Córdoba.

² Departamento de Estadística, Econometría, Investigación Operativa y Organización de Empresas. E.T.S.I.A.M. Universidad de Córdoba. Córdoba.



Asociación Nacional de Criadores
de Ganado Vacuno Selecto de Raza Retinta

Análisis de series de testaje de ganado retinto mediante redes neuronales artificiales: primeros resultados

INTRODUCCIÓN

Valoración de reproductores retintos en el Programa de Selección

Dentro del marco del Esquema Nacional de Selección de la Raza Retinta, aprobado mediante Resolución Comunicada con fecha 29 de Enero de 1.992, la valoración de los animales candidatos a formar parte del grupo de reproductores reviste una especial importancia. Esta evaluación se lleva a cabo en tres fases (Rodero *et al.*, 1999), una primera en la propia finca y que incluye una valoración sobre su crecimiento desde el nacimiento hasta el destete, una segunda en centros de testaje en los que se controla a los machos jóvenes que han superado la fase anterior y una última en la que se valoran por la descendencia a estos reproductores.

De las tres fases reseñadas, vamos a ocuparnos en este artículo de la segunda, en la que animales con cierto mérito demostrado en la fase anterior y que superan unos requisitos adicionales —entre los que destacan los relativos a su ascendencia— son sometidos durante un tiempo a unas condiciones ambientales homogéneas, con la esperanza de que los datos así rendidos estén menos influidos por fuentes de variabilidad ajenas a la propia valía del candidato. Estos datos serán integrados mediante un índice de selección que ofrecerá en una escala numérica ordenada las valoraciones como reproductores de los animales en cuestión.

Ahora bien, el problema que se plantea de forma inmediata tras el escenario descrito no es otro que el de la dificultad de comparación de los animales de unas

series frente a los de otras: en efecto, la homogeneidad que se impone *dentro* de cada serie se traduce de forma inmediata en una heterogeneidad *entre* series que impide la comparación entre los animales que fueron sometidos a control en diferentes tandas.

El objetivo de este trabajo es el de diseñar una metodología, basada en las Redes Neuronales Artificiales que ayude a la depuración de los datos frente a los condicionantes ambientales bajo los que se haya desarrollado cada serie, de forma que el dato corregido que se obtenga resulte más próximo al verdadero valor de cada individuo, por haber sido explicada mediante el modelo en cuestión una parte mayor de su variabilidad.

Redes Neuronales Artificiales

La aparición y progresivo perfeccionamiento de los ordenadores en este último medio siglo ha permitido el abordaje de problemas cada vez más complejos; uno de los frentes más ambiciosos ha sido desde un principio el de la Inteligencia Artificial —intento de emulación de capacidades cognitivas de alto nivel, típicamente humanas como son el razonamiento o el aprendizaje— por parte de las máquinas. Dentro de los paradigmas en los que se divide el movimiento de la Inteligencia Artificial podemos ubicar el conexionismo o Tecnología de Redes Neuronales Artificiales.

Como su propio nombre sugiere, las redes neuronales intentan emular la arquitectura masivamente paralela y fuertemente interconectada de los cerebros biológicos para obtener resultados en los que otros abordajes no rin-

den adecuadamente (Lippmann, 1987). Tal es el caso del tratamiento de datos difusos o con ruido, situación en la que podemos encuadrar sin duda la gestión de los registros fenotípicos de animales. La propiedad fundamental que permite el tratamiento de problemas reales mediante redes neuronales es la de aprendizaje, esto es, la posibilidad de que el algoritmo se perfeccione a sí mismo gracias a los ejemplos que se le proponen y sea capaz de extraer consecuencias útiles de ese proceso.

El empleo de esta tecnología en diversos campos de la ciencia y la técnica ha sido abundante: por hacer una sucinta referencia a los de interés agrario, Yang y cols. (Yang *et al.*, 2000; Lacroix, *et al.* 1997) han hecho uso de esta técnica para clasificar vacas lecheras en mastíticas o sanas en función de datos de la propia granja, en la estimación de la producción lechera, etc. Asimismo se encuentran ejemplos de aplicación de las redes a la clasificación de productos agrícolas (Jayas *et al.*, 2000) o a la de vegetales en variedades en función de datos biométricos procedentes de diversas partes de la planta.

En un afán de resumir, podemos decir que las redes funcionan básicamente de dos formas: clasificando casos propuestos en categorías o estableciendo regresiones que permitan elaborar estimaciones o descubrir alguna dependencia funcional entre variables. Para la consecución de tales fines, la red dispone de unos elementos de proceso (que emulan a las neuronas) dispuestos por capas y masivamente interconectados (donde cada unión recuerda a una sinapsis y viene afectada por un coeficiente o peso); el ajuste de los pesos de las interconexiones mediante un algoritmo de aprendizaje al efecto es lo que permite a la red converger hacia su objetivo. Según la forma en que se desarrolle el aprendizaje se hablará de supervisado o no supervisado, entre otros. En nuestro caso, la red se entrena mediante un algoritmo de aprendizaje supervisado que muestra al programa de qué forma se relacionan las entradas con la salida; es de esperar que, finalizado el proceso de aprendizaje, el programa perfeccionado sea capaz de extraer las reglas que subyacen al proceso aprehendido y esté en condiciones de generalizar ese conocimiento y producir estimaciones de salida frente a nuevos datos de entrada que se le faciliten.

MATERIALES Y MÉTODOS

Fuente de datos

Los datos para este estudio proceden de los controles realizados sobre los animales objeto de estudio en el CENSYRA de Badajoz y el CEAG de Jerez de la Frontera (Cádiz), que son integrados y tratados, en su caso, en el Departamento de Genética de la Universidad de Córdoba, dentro de la participación que a cada una de las citadas Instituciones corresponde en el ya citado Esquema de Selección.

Las variables que se han incorporado a este estudio son las siguientes:

1. Variables independientes, regresores o factores ambientales:

- Centro en el que se lleva a cabo la serie de testaje.
- Estación del año en la que tiene su inicio la serie.
- Duración de la serie.

2. Variables dependientes, a explicar o medidas fenotípicas:

- Peso a los doce meses de cada animal (este dato no es una medida directa sobre el animal, sino el fruto de una aproximación local usando una regresión).
- Ganancia media diaria durante el testaje.
- Calificación morfológica recibida por el animal.

Las series que en este trabajo se consideran van de la 1ª a la 29ª, con un total de 679 animales, tras depurar aquellos que fueron descalificados o sus datos eran incompletos. Esta series tienen su inicio en 1975 (Álvarez *et al.*, 1999) y se extienden hasta 1998. Las primeras series tuvieron una duración cercana a los siete meses y se desarrollaron en el CENSYRA de Badajoz; con la implantación del ya mencionado Esquema Nacional, la duración de las series se redujo al entorno de los cinco meses y se incorporó el Centro de Jerez de la Frontera (Cádiz).

Con ocasión de la puesta en funcionamiento del Esquema de Selección, se modificaron igualmente los caracteres objeto de medida sobre los animales. Las dieciocho primeras series recababan datos relativos al peso a los doce meses, la ganancia media diaria, el índice de conversión y la evaluación morfológica; a partir de la decimonovena se dejó de determinar el índice de conversión y se incluyó la medida de la circunferencia escrotal a los doce meses.

El conjunto de animales se ha segregado al azar en tres subconjuntos, uno llamado de entrenamiento, sobre el que la red intenta materializar el aprendizaje, uno de prueba, que sirve al programa para orientar sus progresos y uno de producción, ajeno completamente al proceso de entrenamiento de la red y que nos servirá para medir la capacidad de generalización de la misma. En nuestro caso, 409 toros fueron adscritos al primer conjunto y 135 a cada uno de los dos segundos (20% del total de animales).

En este experimento hemos usado una red de retropropagación en la que la capa oculta está escindida en tres subcapas para que cada una de ellas pueda captar rasgos característicos del problema en cuestión; esta capacidad de discriminación se consigue mediante el uso de distintas funciones de transferencia. La implementación de esta red se ha llevado a cabo con el programa NeuroShell 2^o versión 4.0 (Ward Systems Group Inc.)

El número de neuronas en la capa de entrada es de cinco (una para codificar el centro de testaje, tres para codificar las épocas del año y otra, continua, para indicar la duración en días de la tanda); veinticuatro neuronas en la capa oculta (ocho para cada subcapa) y tres neuronas en la capa de salida, una por cada variable dependiente considerada en el problema.

RESULTADOS Y DISCUSIÓN

Los parámetros obtenidos sobre el conjunto de producción son los que se insertan:

Tabla 1
Medidas de ajuste obtenidas en las que se aprecia la proporción de varianza atribuible a los factores considerados

Variable de salida	Peso 12M	Ganancia	Puntos
R ² :	0.2828	0.3604	0.1659
r ² :	0.3180	0.4047	0.1769
Error cuad. medio:	2378.968	0.034	27.064
Error abs. medio:	37.101	0.133	3.983
Error abs. mínimo:	0.171	0	0.007
Error abs. máximo:	170.571	0.900	14.656
Coefficiente de correlación:	0.5640	0.6361	0.4206

Desde nuestro punto de vista, el dato más importante que puede extraerse de este avance es el comportamiento diferencial que exhiben las tres variables de salida, en términos de proporción de varianza explicada gracias al modelo (atribuible a los factores considerados).

Se observa cómo la variable ganancia se explica en un 36%; esta proporción cae algo para el carácter peso a los doce meses: el hecho de que se trate de un dato no obtenido directamente sobre el animal, sino que a su vez proviene de una aproximación local gracias a una regresión incluye una variabilidad "oculta" que merma el ajuste de nuestro modelo.

La contribución relativa de cada factor en la explicación de la variabilidad se expresa mediante la gráfica siguiente:

Ahora bien, este efecto de caída en la capacidad de explicación es máximo en lo relativo a la valoración morfológica de los animales; si comparamos con la variable ganancia, observamos una merma del 54%, lo que resultaría compatible con los resultados obtenidos por otros autores (Goyache *et al.*, 2001) sobre la calificación morfológica de razas autóctonas españolas en el sentido de revelar una fuerte carga de subjetividad en el sistema de valoración regional por puntos.

En ese sentido, López *et al.* (2000) han acometido la tarea de proveer de herramientas objetivas para la valoración morfológica usando ciertos paradigmas de la inteligencia artificial, entre los que se encuentran algoritmos que aprenden de ejemplos. Si bien la aproximación de estos autores no es coincidente con la de las redes neuronales, parece claro que la interacción de diversos paradigmas de inteligencia artificial puede aportar elementos que reduzcan la objetividad presente en este tipo de medidas.

La aplicación de la red entrenada cuyos datos se han referido genera un conjunto de registros corregidos frente a los factores ambientales susceptible de ser integrado en un único índice de selección.

A continuación se insertan los pesos resultantes del entrenamiento de la red, haciendo significar que las neuronas de la 1 a la 8 en la capa oculta tienen por función de transferencia la gaussiana, las neuronas de la 9 a la 16, la complementaria de la gaussiana y las ocho últimas, la tangente hiperbólica.

Los valores de entrada son ponderados con los coeficientes que figuran en la primera sección de la tabla; a ese producto se le suma el término que figura en la columna "bias". El valor así obtenido es la entrada de cada neurona de la capa oculta; este valor es sometido a la acción de la función de transferencia de la capa oculta (cada subsección de esta capa tiene una función propia). El valor resultante de esta aplicación es la salida de cada neurona de la

Figura 1
Contribución relativa de los distintos factores a la explicación de la varianza.

- 1: Centro donde se lleva a cabo la serie;
- 2: Época del año de entrada en la serie;
- 3: Duración de la serie

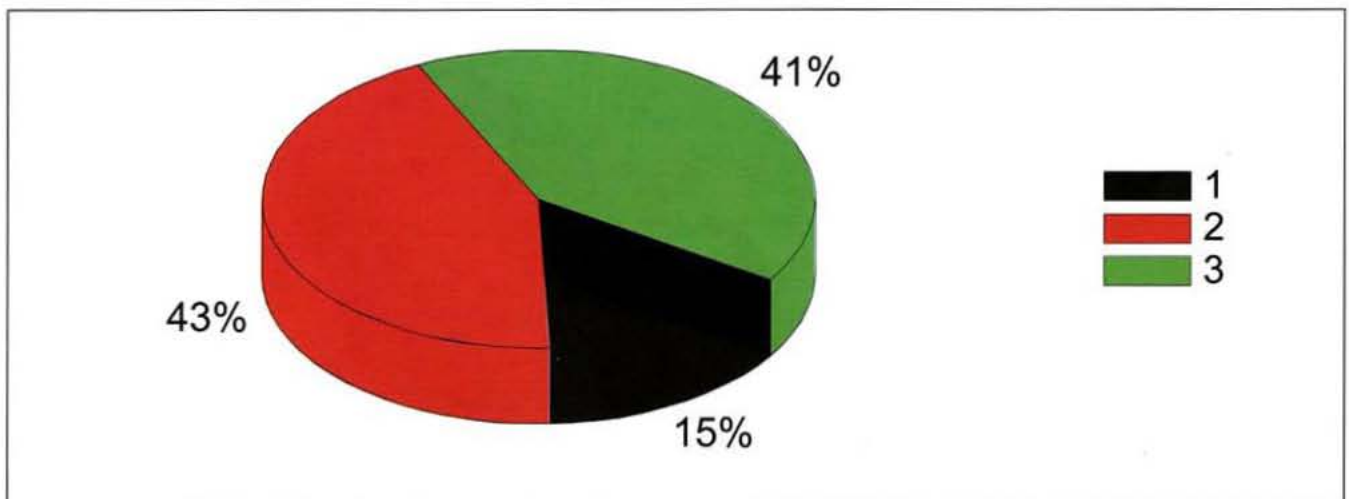


Tabla 2.

Pesos obtenidos del entrenamiento de la red. La primera parte (24 x 6) de la tabla recoge los pesos de las sinapsis entre la capa de entrada y la capa oculta; las secciones restantes (3 x 9), los pesos que van de la capa oculta a cada una de las tres variables de salida. Se observa la presencia de un término de tendencia ("bias"), que equivale a una entrada adicional y fija para cada neurona.

	bias	1	2	3	4	5
1	-0.2006356	0.1623103	-0.1728751	-0.3379133	-0.1571689	-0.1074919
2	0.0280107	0.0739494	-0.1139852	-0.1166181	-0.0025846	0.0603204
3	-0.1864996	-0.0596288	-0.1582987	-0.1328912	-0.1647172	-0.0192429
4	0.7162076	0.0959281	0.0889931	-0.2979317	-0.4151602	2.688062
5	0.0336484	0.1097176	-0.1090831	0.2905021	0.360101	1.4997584
6	-0.1016635	-0.2173771	0.2458461	-0.0400106	0.097031	-0.6949535
7	-0.1153026	-0.0303306	-0.155156	-0.1403776	-0.0958672	-0.0015074
8	0.0909353	-0.2475135	-0.1795268	0.021802	-0.0138389	-0.3990468
9	0.1855039	0.0644815	0.4994175	-0.1684042	-0.4963389	-0.5210152
10	-0.2978778	0.4197453	0.0380007	-0.0444406	0.498493	0.5679848
11	0.0997682	0.1354545	-0.5170639	0.5814889	0.3922971	-0.5518996
12	0.4155947	-0.4996486	-0.2547308	-0.3605446	-0.2753806	-0.9494889
13	-0.6388776	0.0507278	-0.0312903	0.4304729	0.5684724	0.1686991
14	-0.2741046	0.4603396	-0.2498787	0.3932064	0.4940916	0.1497951
15	-0.3482365	0.296592	-0.5602056	0.0720668	0.4366947	-0.3442781
16	0.5207902	-0.210093	0.229798	-0.1964889	-0.5224921	-0.2788061
17	0.0103152	-0.2119582	0.5193267	-0.4144235	-0.6072011	-0.4006554
18	0.5367936	-0.6321443	0.0527107	-0.1215029	-0.6607625	0.1613202
19	0.63848	-0.3402548	-0.444931	-0.1715694	0.3321551	1.9167033
20	0.0294814	-0.6397706	0.3020343	0.0843227	0.0050047	0.6484556
21	-0.3646265	0.71456	-0.0849399	0.1872365	0.0377988	0.6574749
22	0.4581889	0.1435454	0.3657082	-0.0399557	-0.3675084	-0.4319965
23	-0.3413534	-0.0205782	-0.1152202	-0.1902049	0.1184649	-0.4500996
24	0.6021944	-0.3316179	-0.0242792	-0.3906162	-0.493503	-0.5881045

	bias	1	2	3	4	5	6	7	8
1	0.1448239	0.0351942	-0.0658806	-0.2495504	-0.3064341	-0.5049132	-0.4182532	0.0227625	-0.3242426
2	-0.3012992	0.2439563	-0.2433614	0.0538965	-0.9017726	-0.4466926	-0.3360194	0.0494748	-0.2235836
3	-0.0832923	0.0866932	-0.3720088	-0.3674685	0.5484866	0.7550763	0.293196	-0.0967056	0.0108337

	bias	9	10	11	12	13	14	15	16
1	-0.0312741	0.2059438	-0.1354516	0.2569349	-0.0529829	-0.0252391	-0.0697956	0.0993743	0.0457957
2	-0.017714	0.1457183	-0.1374833	0.2339763	0.3732151	-0.0503183	0.1708486	0.1643763	0.0740347
3	-0.0975752	0.0913233	-0.1544964	0.1384401	0.4849969	0.0868684	-0.0313169	0.1540208	-0.0177168

	bias	17	18	19	20	21	22	23	24
1	0.0452099	-0.0003633	-0.0417028	0.4865956	-0.1548607	0.2832711	0.0866278	-0.0050711	0.2342613
2	-0.2528473	0.2642807	-0.240124	0.3311598	-0.1353739	0.2227319	0.1997386	0.1262141	-0.0093097
3	-0.2890713	0.102681	-0.0206516	0.4438269	-0.299562	0.2140551	0.054632	-0.3281816	0.1546011

capa ocular. El proceso se repite desde cada una de las veinticuatro neuronas de la capa oculta a las neuronas de salida; en este caso, la función de activación de todas las neuronas de la capa de salida es la logística o sigmoidea.

CONCLUSIONES

En atención a los resultados y la discusión que sigue a los mismos, parece que dos deben ser los frentes en los que hay que trabajar en un futuro para obtener medidas más precisas sobre la influencia de los efectos ambientales en las variables medidas sobre los candidatos a reproductores:

1. Sustituir la estimación del peso a los doce meses por una medida directa, si es posible, o mejorar la precisión de las estimaciones actuales.
2. Incluir mecanismos que reduzcan la subjetividad de las valoraciones morfológicas de los animales; las técnicas de inteligencia artificial (ya sean bajo el paradigma simbólico o bajo el de redes neuronales) aparecen prometedoras a este respecto.
3. Incorporar nuevos datos de las series y nuevas arquitecturas de red neuronal para obtener modelos más precisos estableciendo una realimentación entre los métodos de evaluación de animales y los diseños de redes que permita redefinir, en su caso, las metodologías empleadas hasta el momento.

BIBLIOGRAFÍA

Álvarez, F., F. Delgado, M. Valera, A. Molina, A. Rodero: Analysis of testing series in the retinto beef

cattle improvement plan. 50th meeting of the European Association for Animal Production. Zúrich, 1999.

Goyache, F., J. P. Gutiérrez e I. Álvarez: Efectos ambientales de influencia sobre la calificación morfológica de reproductoras de raza Asturiana de los Valles. XXXIII Jornadas de Estudio de la Asociación Interprofesional para el Desarrollo Agrario; IX Jornadas sobre Producción Animal. Zaragoza, abril del 2001. Tomo I, pp 130-132.

Jayas, D. S., J. Paliwal y N. S. Vison: Multi-layer neural networks for Image Analysis of Agricultural Products. *Journal of Agricultural Engineering Research*, 77 (2). 2000.

Lacroix, R., F. Salehi, X. Z. Yang y K. M. Wade: Effects of Data Preprocessing on the Performance of Artificial Neural Networks for Dairy Yield Prediction and Cow Culling Classification. *Transactions of the ASAE*, 40 (3). 1997.

Lippmann, R. P.: An introduction to Computing with Neural Nets. *IEEE ASSP Magazine*, Abril 1997.

López, S., F. Goyache, J. R. Quevedo, J. Alonso, J. Ranilla, O. Luaces, A. Bahamonde y J. J. del Coz: Un sistema inteligente para calificar morfológicamente a bovinos de la raza Asturiana de los Valles. *Inteligencia Artificial*, 10, pp5-17. 2000.

Rodero, A., A. Molina, M. Valera, F. Delgado, J. M. Jiménez, F. Avilés y J. M. García.: Situación Actual del Esquema Nacional de Selección de la Raza Retinta. *FEAGAS*, nº 15, Enero-Junio 1999.

Yang, X. Z., R. Lacroix y K. M. Wade: Investigation into the production and conformation traits associated with clinical mastitis using artificial neural networks. *Can. J. Anim. Sci.* 80: pp 414-426. 2000

PÍDALA DONDE VEA ESTE DISTINTIVO



Asociación Nacional de Criadores de Ganado Vacuno Selecto de Raza Retinta

Murcia, 5, 1.º A • 28045 Madrid
Teléf.: 91 468 22 05 • Fax: 91 467 80 00
E-mail: razaretinta@interbook.net

Más de **25** años mejorando la Raza

