Using artificial intelligence to design and implement a morphological assessment system in beef cattle

F. Goyache[1]†, J. J. del Coz[2], J. R. Quevedo[2], S. López[2], J. Alonso[2], J. Ranilla[2], O. Luaces[2], I. Alvarez[1] and A. Bahamonde[2]

[1]*SERIDA-CENSYRA-Somió, C/ Camino de los Claveles 604, E-33203 Gijón (Asturias), Spain*
[2]*Centro de Inteligencia Artificia, Universidad de Oviedo in Gijón, Campus de Viesques. E-33271 Gijón (Asturias), Spain*

†E-mail: felixgg@princast.es

Abstract

In this paper a methodology is developed to improve the design and implementation of a linear morphological system in beef cattle using artificial intelligence. The proposed process involves an iterative mechanism where type traits are successively defined and computationally represented using knowledge engineering methodologies, scored by a set of trained human experts and finally, analysed by means of four reputed machine learning algorithms. The results thus achieved serve as feed back to the next iteration in order to improve the accuracy and efficacy of the proposed assessment system. A sample of 260 conformation records of the Asturiana de los Valles beef cattle breed is shown to illustrate the methodology. Three sources of inconsistency were detected: (a) the existence of different interpretations of the trait's definition, increasing the subjectivity of the assessment; (b) the narrow range of variation of some of the anatomical traits assessed; (c) the inclusion of some complex traits in the assessment system. In this sense, the reopening of the evaluated Asturiana de los Valles assessment system is recommended. In spite of the difficulty of collecting data from live animals, further implications of the artificial intelligence systems on morphological assessment are pointed out.

Keywords: *artificial intelligence, beef cattle, linear type, machine learning.*

Introduction

Morphological assessment is a common activity in many current improvement programmes in cattle. In dairy cattle, moderate genetic correlations have been found between type traits and herd life or production traits (Brotherstone and Hill, 1991; Short and Lawlor, 1992). Consequently, type traits are considered to be good indirect predictors of stayability and dairy yield traits (Brotherstone, 1994; Vukasinovic *et al.*, 1997). Beef cattle improvement programs usually include muscular development and skeletal development assessment in calves (Shi *et al.*, 1993), though there is not much information on the use of type traits in beef cattle. In beef production, type traits are expected to be useful in evaluating the animals as meat producers. However, there is a lack of scientific consensus on the possibilities of improving productive traits as a result of selection of conformation traits in live cattle (Berg and Butterfield, 1979). The characteristics of classical methodologies for the assessment of type traits are probably at the root of this possible lack of reliable results. Scoring type traits on a 'desirable' scale has been reported as an inadequate methodology for carrying out morphological assessment (Vukasinovic *et al.*, 1997). The use of these subjective criteria instead of a linear scale revealing anatomical traits might result in a loss of biological relationships among traits. In this sense, the visual appearance of beef cattle could be affected by many optical tricks that make it difficult to select beef characteristics in live bovines in order to increase meat production ability.

To solve these problems, the International Committee on Animal Recording (1995) recommends the standardization of a linear conformation recording system. The linear methodology is expected to be a more accurate tool for recovering biological relationships among traits. In the linear

morphological assessment system, all traits must be scored linearly from one biological extreme to the other, irrespective of an animal's sex, age or type. However, the design of linear systems could allow different recording scales according to the sex or age of the assessed animal (Associazione Nazionale Allevatori Bovini Italiani Carne (ANABIC), 1997), that might introduce undesirable sources of subjectivity.

In addition, the consistency of linear morphological assessment and its interpretation is strongly influenced by the so-called classifier effect (Fernández *et al.*, 1995; de la Fuente *et al.*, 1996). The assessment of linear type traits can also be biased by the influence of herd of origin. Regardless of the applied assessment system, the general appearance, size, age and sex of the animal affect the classifier's accuracy. Moreover, the definition of morphological traits assessed using the linear methodology might not be sufficiently accurate, especially in complex traits such as compactness or angularity. In spite of intense training, different classifiers could make various interpretations when scoring the same trait. All these factors can affect the field performance of type trait recording. In fact, type trait records could exhibit a lack of linearity despite the fact that they are expected to show a linear variation (López *et al.*, 2000).

Linko (1998) summarizes the advantages of AI systems as: (a) they may help when expert advice is needed but an expert may not be available; (b) they are independent of human errors or moods; (c) they can help to verify a human expert's opinion; (d) they are available 24 h/day; (e) they can operate in risky situations; (f) they can act quickly on the basis of huge databanks and knowledge banks; (g) they can use natural language, and do not require complex mathematical expressions. All these reasons lead to the application of AI in animal production. Most projects involving AI and animal production are directed towards the implementation of expert systems useful for helping breeders to take management decisions (i.e. culling decisions) or to carry out diagnosis of pathological processes (Pietersma *et al.*, 1999; Yang *et al.*, 1999). These systems deal with knowledge extracted from an expert and which is then represented in such a way that it can be run on a computer. However, AI can go further and help us to discover or learn this knowledge by just observing the behaviour of a breeder or an expert qualifier when they are taking some decision affecting the performance of cattle.

The aim of this paper is to propose the use of artificial intelligence (AI) to develop morphological

assessment systems in cattle. The rôle of machine learning (ML) algorithms is to clone, in a systematic and plausible way, the behaviour of (in this case) experts in morphological assessment. Thus, the difficulty of a concept to be learned and performed can be estimated as a measure of the coherence of the recorded assessments. The analysis of the prediction errors from several ML systems and their comparison with the results obtained from regression is useful for evaluating the coherence of field data. In addition, AI is expected to detect whether the inconsistency of the recorded data originates in a deficient definition of the assessed trait or from a misunderstanding of classifiers. A sample of Asturiana de los Valles beef cattle is used to show the reliability of the proposed methodology.

Material and methods
Description of AI algorithms
Artificial intelligence can be seen as that part of computer sciences that tries to simulate processes that in humans would be described as intelligent behaviour. Thus, machine learning is one of the central topics of AI, since a feature usually attached to intelligence is that of the ability to learn from the environment. From a general point of view, ML algorithms synthesize knowledge from an unorganized source in such a way that their outputs are in fact computer routines able to accomplish useful tasks.

The most common input data for learning functions are sets of examples of wise actions performed in the past by human experts collected in so-called training sets. These examples are described by a set of attributes with a singular attribute called 'class' of the example. In this context, the outputs of these functions are other functions or, in general, pieces of knowledge able to compute (predict) the class of unseen cases (examples without a known class). ML algorithms are conceptually near to regressions, but prediction functions obtained from ML algorithms can be more sophisticated. Typically, when the class to be predicted is a continuous number, a finite set of rules specifying in which case to apply a linear function can be obtained. Thus, instead of having only one function, as happens with regression, some ML algorithms provide a set of linear functions that can approximate any kind of regular function not necessarily linear. See textbooks like Nilsson (1998), Rich and Knight (1991), Quinlan (1993a) or Michalski *et al.* (1998) for a detailed explanation of ML principles and methods.

Three different types of ML algorithms were used to judge the quality of the morphological assessment

Table 1 *Features of machine learning (ML) algorithms producing sets of linear functions used in this paper*

ML system	Kind of output	Evaluation mechanism
Cubist	Set of rules producing linear functions	Exact (crisp)
M5'	Decision tree with linear functions on their leaves	Exact (crisp)
SAFE	Set of rules producing linear functions	Nearest neighbour (fuzzy)

system for Asturiana de los Valles bovines. The outputs of these kinds of algorithms were as follows.

(a) Decision trees formed by nodes with the conditions that should be accomplished, generating linear functions used to predict numeric values.

(b) A set of rules whose premises or conditions are expressed by logic clauses about attribute values and whose results are the linear functions that should be used to compute the final result.

(c) A set of prototypes chosen from the examples in the training set. These algorithms, which are called instance-based learners, memorize prototypes and use them to predict the result in similar situations.

At the same time, two prediction mechanisms are applied in ML systems to classify new examples: exact agreement (or crisp evaluation), and the nearest neighbour principle (or fuzzy evaluation). In the exact agreement method, all of the conditions must be true to make use of its associated prediction function or value. However, if these requirements are relaxed, fuzzy evaluators should be used. In this latter case, the rule selected to evaluate a new case need not necessarily fulfil all its conditions completely, but rather its application conditions should be more similar to the new case than any other rule. Thus the metric used to measure the similarity between two instances is very important. The fuzzy evaluation mechanism is typical of the systems that remember prototypes.

In the field of ML, there are many algorithms of each type. To achieve the goals of the present paper, we used algorithms that learn rules, such as Cubist (Quinlan, 2000) and SAFE (Quevedo and Bahamonde, 1999), and decision trees, like M5' (Quinlan, 1993b; Wang and Witten, 1997). Table 1 summarizes the main characteristics of these ML systems. Figure 1 gives an example of the different outputs obtained

---

    Simple regression
GM = − 3·13787 + 0·0882 V2 + 1·41 V3   for all cases

    SAFE's output
GM = + 4·96615* V3 + 0·657081* V1 + 0·0938899* V2 −11·0308   if V3 Ɛ [1·43, 1·65] and V1 Ɛ [1·5, 3·17]
GM = + 0·523501* V1 + 3·53035   if V2 Ɛ [62·57, 68·7] and V1 Ɛ [1·96, 4·5] and V3 Ɛ [1·58, 2·2]
GM = + 5·51948   if V2 Ɛ [69·33, 75·26] and V3 Ɛ [1·51, 1·8]
GM = + 1·11282* V3 + 0·0545742* V2 + 0·268809   if V2 Ɛ [62·57, 87·9] and V1 Ɛ [3·66, 5·41] and V3 Ɛ [1·21, 2·21]

    M5' decision tree
V2 ≤ 79·74
 │ V1 ≤ 3·075
 │ │ V2 ≤ 64·855 LM1 (6/13·6%)
 │ │ V2 > 64·855
 │ │ │ V1 ≤ 2·33 LM2 (3/8·41%)   LM1: GM = − 2·49 + 0·342* V1 + 0·0786* V2 + 0·765* V3
 │ │ │ V1 > 2·33 LM3 (7/28·1%)   LM2: GM = − 1·37 + 0·254* V1 + 0·0685* V2 + 0·702* V3
 │ V1 ≥ 3·075 LM4 (29/56%)   LM3: GM = − 1·54 + 0·254* V1 + 0·0685* V2 + 0·814* V3
V2 > 79·74   LM4: GM = 2·04 + 0·080* V1 + 0·0391* V2 + 0·37* V3
 │ V2 ≤ 84·73 LM5 (8/46·2%)   LM5: GM = 14·1 − 0·101* V2 + 0·634* V3
 │ V2 > 84·73   LM6: GM = 7·21 − 0·119* V1 − 0·0084* V2 + 0·792* V3
 │ │ V1 ≤ 5·695 LM6 (8/18·2%)   LM7: GM = 6·32 − 0·144* V1 + 0·0061* V2 + 0·634* V3
 │ │ V1 > 5·695 LM7 (4/22·5%)

Figure 1 Formulae obtained by simple regression, by SAFE and by M5' to assess general muscularity (GM) of beef cattle. The variables (V1, V2, V3) used are defined in Table 3. SAFE provides 4 linear maps, while regression returns only one. M5' provides a decision tree whose leaves indicate a linear function that must be used to compute an assessment; in this case, seven different linear functions (LM 1 to 7) are used.

by simple regression, by SAFE and by M5' to assess a complex trait included in the morphological assessment system (general muscularity) used by the Asturiana de los Valles Breeders Association (ASEAVA). We emphasize the fact that SAFE provides four linear maps, while regression, returns only one.

In addition to these ML algorithms, we used a system of a quite different kind, BETS (del Coz *et al.*, 1999, del Coz and Bahamonde, 1999; Bahamonde *et al.,* 1997; Luaces *et al.,* 1999), which is a system that remembers prototypes. The idea is to have more reliable evidence about how difficult it is to learn how to assess type traits. Consequently, the accuracy of the results from the assessment of beef cattle could not be affected by bias due to the learning method.

BETS (best examples in training sets) is a learning system derived from the nearest neighbour algorithm (Cover and Hart 1967). It selects the most representative instances of the training set and considers that these examples are useful for classifying the elements of the domain. To predict the numeric value of an unseen case *q*, BETS searches for the two nearest remembered instances and then calculates the distance between *q* and these two instances. The final result is an interpolation, inversely related to the distance from *q*, of the values attached to the nearest neighbours. The measure used by BETS to compute the similarity between a recorded instance *x* and a case *q* is:

$$d(x,q) = \sqrt{\sum_{a \varepsilon A} w(a,x) \cdot (x_a - q_a)^2} \qquad (1).$$

Here *A* is the set of attributes, x_a the value of instance *x* in attribute *a*, and *w* reflects the importance of attribute *a* in resolving the problem. This last factor (*w*) is most important in the learning strategy of BETS, because it represents the relevance of each attribute, in the surroundings of recorded cases.

Of course there are other alternative families of ML algorithms that have not been used in this paper. This is the case of artificial neural networks (ANNs). ANNs do not return explicit functions as does BETS. In this sense, ANNs could be used to evaluate the difficulty of learning how to assess cattle morphologically, but ANNs need the previous definition of the layout and other parameters prior to performing the learning process. Under these conditions, it might have been difficult to know whether the results could be worse than expected because of deficiencies in the customization process instead of in the learnability of training sets. These characteristics led us to reject the use of ANNs.

Knowledge engineering and data collection
Knowledge implies understanding. The process of constructing a knowledge-based system is called knowledge engineering. This concept entails the acquisition, representation, implementation and processing of knowledge to solve practical problems. Knowledge representation is a central subject in AI (Linko, 1998; Nilsson, 1998; Rich and Knight, 1991) and the key to a successful ML system. Knowledge and procedures normally used by human experts in solving problems within a well defined domain are modelled and translated into a formal system of symbols that can be stored on a computer. Clearly, knowledge representation should serve to solve the problem by incorporating appropriate data and knowledge into the system. In the case of ML, knowledge is handled both as an input (although unstructured) and as the output in the form of a computable prediction function.

A total of 260 type assessment records from four trained expert classifiers of the ASEAVA were available. These records were collected from 65 Asturiana de los Valles animals. Twenty-one male (seven younger than 2 years and 14 older than 3 years) and 44 female (11 heifers and 33 multiparous cows). These 260 records made up the training set containing two elements: the assessments of the experts and the zoometric measurements of the animals. The animals involved in the experiment were selected to include the entire observable morphological variability in the Asturiana de los Valles breed, ranging from the animals classified as excellent to poorly conformed, in order to avoid sampling bias. Consequently, the training set can be considered as a representative collection of examples useful for learning to assess all kinds of cattle in the breed.

Four expert classifiers assessed 21 type traits included in the ASEAVA's linear type traits assessment system (Goyache *et al.,* 1999) (Table 2) for evaluating the animal's muscular and skeletal development. All assessed traits are scored on a linear scale from 1 to 9. Although most of the assessed traits are indicators of single anatomical traits, three of them (beef appearance, general muscularity and compactness) involve several. Once the linear system had been defined, the experts were trained on a set of 50 animals (18 males and 32 females) to avoid individual differences in cattle assessment. This first set of animals was not included in the analysis of ML algorithms. Subsequently, the experts individually assessed the 65 animals of the training set.

Table 2 *Muscular development and skeletal development traits information for the analysed database*

Trait	Abbreviation	Score 1	Score 9	Mean	s.d.
Muscular development					
Shoulders convexity	SC	Concave	Convex	6·15	1·31
Loins development	LD	Thin and short	Large and long	5·89	1·40
Rump convexity	RC	Meagre	Very convex	5·87	1·36
Buttocks length	BL	Short	Long	6·32	1·13
Buttocks convexity	BC	Concave	Convex	6·12	1·24
Thigh thickness	TT	Thin	Large	5·74	1·44
Beef appearance	BA	Angular	Coarse	6·18	1·45
General muscularity	GM	Poorly muscled	Highly muscled	6·02	1·38
Compactness	CO	Not compact	Compact	6·11	1·48
Skeletal development					
Chest width	CW	Narrow	Wide	5·85	1·37
Height at withers	HE	Small	Tall	5·75	1·29
Chest depth	CD	Shallow	Deep	6·07	1·26
Depth of belly	DB	Deeper	Hound belly	5·91	1·13
Body length	BL	Short	Long	6·08	1·24
Rump length (from hips to pins) (tuber coxae-tuber ischii)	RL	Short	Long	5·97	1·18
Rump angle (from hips to pins)	RA	Low pins	High pins	6·20	0·84
Rump slope rear view	RS	Steep	Flat	6·33	1·18
Pins width (tuber ischii)	PW	Narrow	Wide	5·36	1·24
Haunch width	HW	Narrow	Wide	6·02	1·21
Lumbodorsal line	LL	Concave	Very convex	6·28	0·80
Head length	HL	Short	Long	5·37	0·85

The average score given by the experts gave the class or learning target for each animal and each type trait. This datum needs to be endowed with quantitative attributes able to numerically explain the score given to the trait. Thus, three digital images (frontal, lateral and rear) were obtained for each evaluated animal. These images were then processed by marking 63 key points to calculate the zoometric measurements used in the assessment process (Figure 2). Single anatomical traits were measured by means of length (i.e. HE = distance between L1 and L3), angle with respect to ground line (i.e. RA = angle with L12-L19 line) or relationships between circumferences determined by subsets of three points (i.e. SC is related to circumference F12-F2-F3). Complex traits were represented by means of geometrical relations
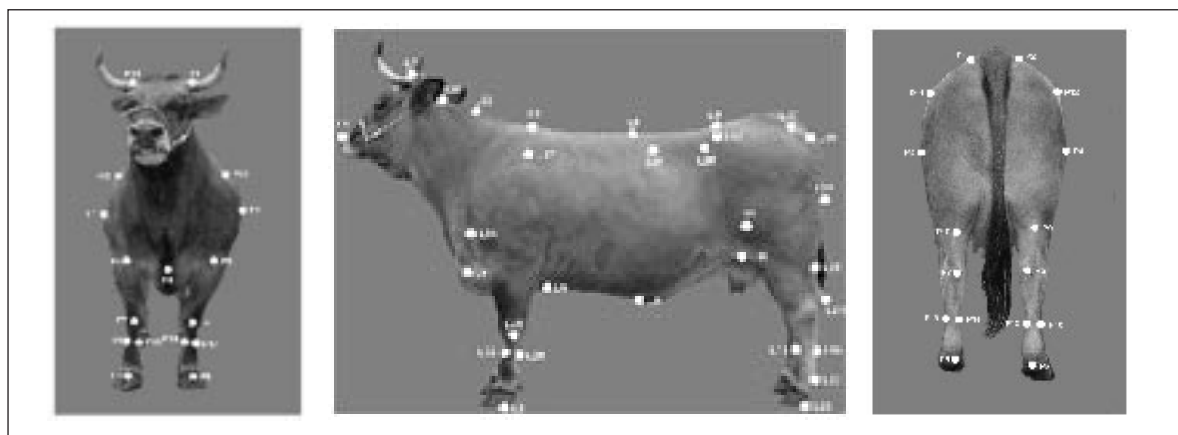

Figure 2 Pictures (front, lateral and rear) obtained for each assessed animal with the key points marked on.

Table 3 *A fragment of the training set used to learn how to assess the general muscularity of beef cattle*

Attributes quantifying general muscularity that is, degree of roundness			Learning target or class: experts' average assessment	Assessment learned by SAFE from all the other animals	Absolute difference
V1: buttocks in lateral view	V2: buttocks in rear view	V3: chest in front view			
4·83	85·41	2·02	8·39	7·15	1·24
5·13	85·44	2·36	7·99	7·59	0·40
3·36	74·33	1·79	5·99	5·94	0·05
2·81	70·41	1·90	5·99	5·64	0·35
2·36	72·80	1·62	6·19	5·59	0·60
5·41	82·08	1·63	6·25	6·27	0·02
4·50	67·54	1·58	6·39	5·99	0·40
1·84	50·17	1·85	3·00	3·73	0·73
2·83	56·31	1·65	4·00	4·42	0·42
2·65	63·28	1·75	4·00	4·76	0·76

† Notice that the fragment only contains 10 entries, the complete training set has 65 entries. The first three columns describe the attributes or zoometric quantification of animal features relevant to computing the target property: the experts' average assessment about general muscularity. The last two columns describe individual accuracy estimated by applying to each animal what was learned from a training set formed by all the other 64 entries.

between different distances and circumferences. The precision of the attained digital zoometric measurements was successfully tested by comparing them with the zoometry obtained using a Lydtin stick. Additionally, we did not observe any bias due to the subjectivity of the operator who marked the points on the images.

For each trait considered, each scoring event was represented by means of a vector of properties or attribute-value pairs according to the description given to our assessment experts. These attribute-value pairs were zoometric measurements of the animal's body related to the trait that we are trying to learn to assess. With the goal of designing the appropriate representations of the different type traits, the human experts involved described the type traits. These descriptions, together with the predetermined definition of the traits, were discussed in working-group sessions where computer representations were drawn for every trait.

In addition, the score given by our experts to each animal in each type trait was recorded. Table 3 presents some of the training examples to illustrate this phase of the methodology, reporting a piece of
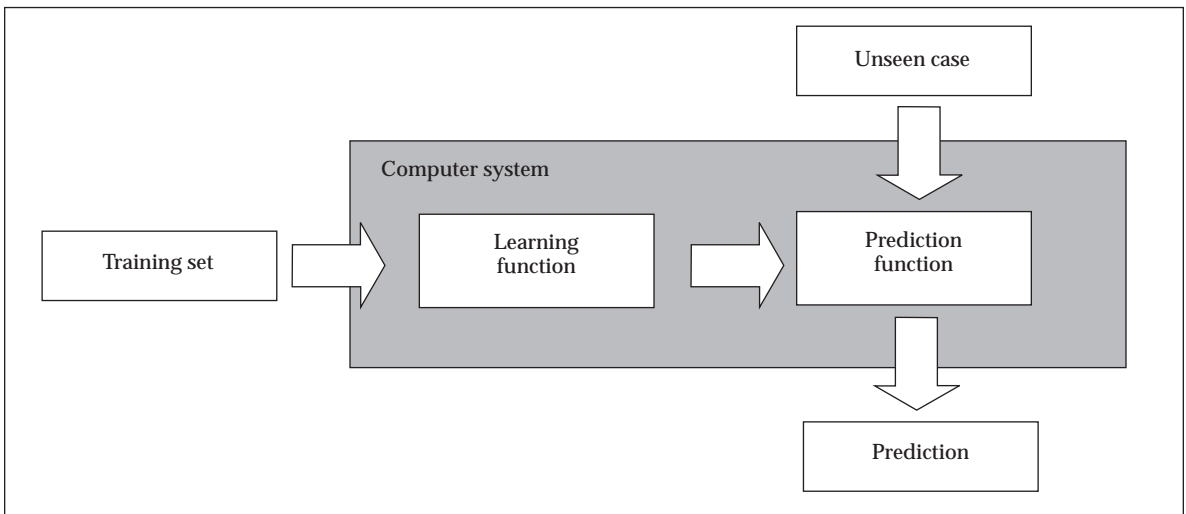

Figure 3 The rôle of machine learning algorithms in knowledge management.

the data collected to learn the general muscularity trait. In order to represent this concept, three attributes describing the lack of angularity and degree of convexity of different parts of the animal's body were used:buttocks in lateral and rear view, and chest in front view. The learning target was the average score given by our experts.

Once the training sets have been compiled, learning algorithms extract the knowledge necessary to predict the assessments of unseen animals according to a plausible model that generalizes the way in which our average expert accomplishes his task. Figure 3 sketches the process followed.

For each one of the 21 type traits considered, and starting from their associated training sets, we conducted a number of computations to determine the difficulty of each assessing task and its degree of linearity. The precision of the ML algorithms is usually estimated by computing the difference between the output of the synthesized prediction function for each animal, and the actual value of the target class (experts' average assessment) using the ML method called leaving-one-out. The leaving-one-out method estimates the accuracy of ML algorithms, separating training and testing sets. For each example in each data set, the AI system learned to assess from all of the other available examples. The average of the differences thus computed is a faithful estimation of the accuracy of the ML algorithm acting on the whole training set when we apply the prediction functions to unseen cases. Subsequently, it uses what has just been learned to compute the score

of each example and compares it with the actual score given by our human experts. Therefore, 1365 (65×21) learning experiments were carried out for each ML algorithm, one for each example and type trait. Table 3 shows a fragment of the results achieved for the general muscularity trait.

Methodology design
Figure 4 sketches the feedback mechanism used to devise a robust and reliable assessment system for beef cattle. This process is quite general practice. The methodology applied by experts is analysed to serve as feedback for the design and implementation of the methodology. The novelty lies in the use of AI techniques. The overall strategy applied in this paper, the design and implementation of a morphological assessment system, involved: a) definition of traits to be considered in the morphological assessment system; b) selection and training of human experts to assess the animals; c) computational representation of the traits under consideration; and d) analysis of the results.

Of course, the scores obtained from experts can be analysed by traditional methods. The coherence and linearity of the experts' scores can be computed by means of correlation coefficients. However, it is not easy using the correlation coefficient to define a threshold to discriminate traits that are rationally evaluated by experts and which should be redefined or even eliminated from the assessment system. A set of points that defines two or more similar linear partial functions can be forced to be understood as a unique linear function if this is the only tool that one
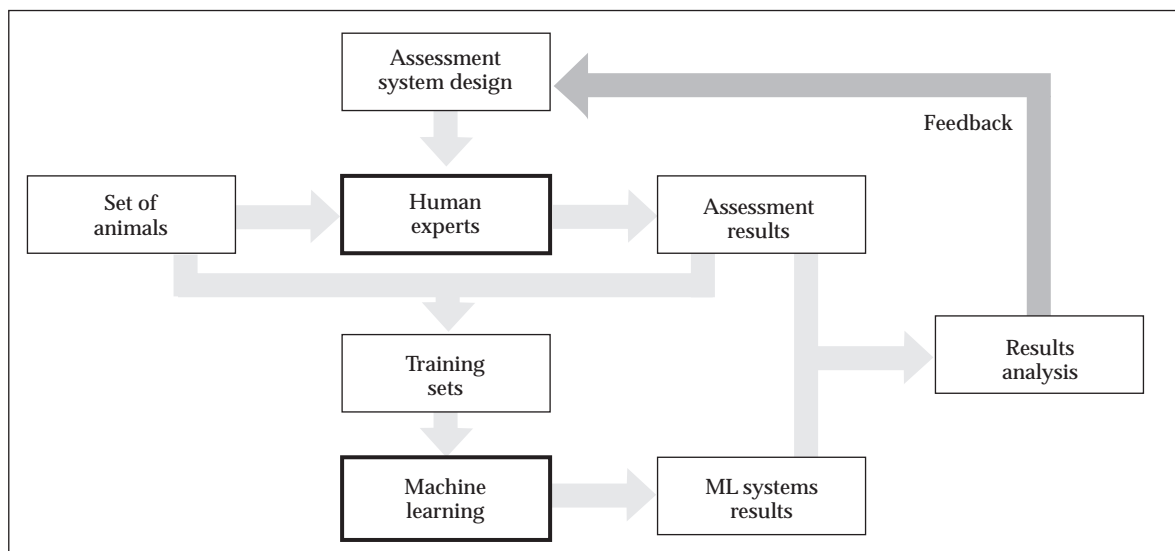

Figure 4 Feedback mechanism to design and implement a morphological assessment system in beef cattle using machine learning algorithms.
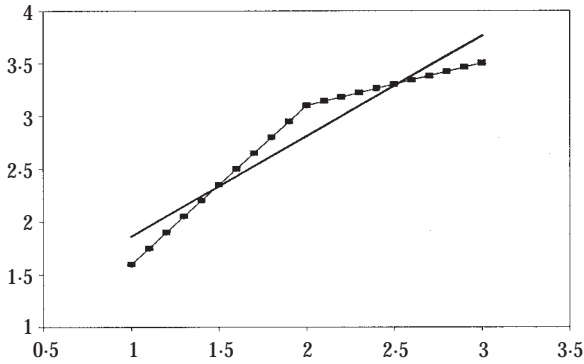
Figure 5 Example of a situation in which a set of points can be highly correlated even though their relationship is non-linear. The variables represented on these axes are highly correlated (0·96), but their relationship is given by a function that is just partially defined as linear: for values of the horizontal axis lower or higher than 2.

formed by the partially linear segments was increased.

The goal of ML is to extract knowledge required to perform a useful task. ML input is not an explicit set of behaviour rules as occurs with expert systems. ML algorithms start from a set of wise actions performed in the past by human experts, collected in a so-called training set of examples. The methodology proposed in this paper, aims at finding computable ways to attach assessment scores to animal measurements (linear) for each trait by means of a set of well reputed ML induction algorithms. The output of these ML algorithms is conceptually of the same kind as regressions. For each trait, they return a procedure to compute morphological assessments once the measurements involved (our training sets) are provided. The predictions functions induced by ML algorithms are a kind of mechanical assessor. The analysis of the results obtained by experts and AI systems will provide the feedback needed to improve the implementation of the morphological assessment system. The core points in the analysis of results are the coherence of the human experts and the degree of linearity of type trait scores. If ML algorithms compute the expert scores as linear regressions, all the stages of definition and

has to explain the behaviour of numerical relationships. Figure 5 depicts one such situation where the correlation coefficient of the whole set of points is very high (0·96) but their relationship is given by a function that is not linear. Of course, lower correlations could be obtained if the angle

Table 4 *Average absolute errors and number of functions (or prototypes) for four machine learning algorithms in beef cattle type traits assessment*

Trait	Cubist		Safe		M5'		BETS		Correlation
	Error	Functions	Error	Functions	Error	Functions	Error	Prototype	
SC	0·63	1·93	0·62	1·29	0·64	1·20	0·71	12·8	0·86
LD	0·98	1·67	0·97	1·97	0·88	4·11	0·93	11·9	0·70
RC	0·87	1·00	0·87	1·23	0·87	1·00	1·07	9·7	0·64
BL	0·62	1·06	0·60	2·40	0·59	4·62	0·78	10·8	0·84
BC	0·85	2·12	0·77	1·75	0·75	4·94	0·63	9·8	0·64
TT	0·58	1·00	0·65	1·07	0·56	1·00	0·79	12·3	0·90
BA	0·74	3·40	0·72	3·37	0·79	5·38	0·60	15·2	0·83
GM	0·84	3·09	0·83	4·17	0·78	5·55	0·86	10·8	0·73
CO	0·71	1·15	0·76	2·71	0·77	3·03	0·87	13·1	0·82
CW	0·56	1·00	0·58	3·22	0·56	1·00	0·66	12·8	0·80
HE	0·46	1·00	0·45	2·21	0·46	1·00	0·54	12·8	0·88
CD	0·50	1·00	0·50	1·05	0·50	1·00	0·55	14·6	0·85
DB	0·52	1·00	0·52	2·09	0·52	1·00	0·58	14·2	0·79
BL	0·54	1·00	0·55	1·26	0·54	1·00	0·71	13·6	0·86
RL	0·63	1·00	0·63	1·00	0·63	1·00	0·72	14·2	0·68
RA	0·46	1·00	0·46	1·00	0·46	1·00	0·50	15·1	0·64
RS	0·64	1·00	0·66	1·40	0·64	1·00	0·69	14·6	0·61
PW	0·83	2·00	0·85	1·02	0·80	2·00	0·89	15·7	0·57
HW	0·61	1·00	0·61	2·83	0·60	1·00	0·76	14·2	0·80
LL	0·49	1·75	0·53	1·54	0·46	2·97	0·46	15·2	0·65
HL	0·50	1·98	0·51	1·00	0·48	3·12	0·61	9·5	0·61
Average	0·64	1·57	0·64	1·94	0·64	2·36	0·71	12·7	

† The errors were computed as the deviation from the human experts' average assessment. Additionally, Pearson's correlation between human experts' average assessments and individual qualifier's assessments are shown.

implementation of the linear assessment system would have been successfully accomplished. However, the linearity of a score should be proven with respect to the representations built at the same time as type traits were defined. Therefore, the linearity of a type scoring will depend on the computational representation, and of course, on the linear appreciation of the scores in the mind of the experts.

Results and discussion

Table 4 describes the results obtained from the computational experiments carried out. The errors were computed as deviations from the experts' average assessment. Additionally, the final column gives Pearson's correlation between the experts' average assessments and the attribute values used to build training examples, since the latter are supposed to represent, in a computational sense, the amounts to be valued.

All of the correlations obtained were significant, ranging from 0·57 to 0·90. There were no clear differences between traits with respect to correlation values. The traits ranging from 0·80 to 0·90 were (as defined in Table 2) SC, BL, TT, BA, CO, CW, HE, CD, BL and HW. Half of these are considered as muscular development traits and the rest as skeletal development traits. Correlation values lower than 0·70 were included mostly in the skeletal development traits.

Surprisingly, complex traits like BA and CO show correlations higher than 0·80. *A priori*, such traits are supposed to present non-linear behaviour despite the linearity of recording (López *et al.*, 2000). Expert classifiers seem to be capable of taking into account very complex information that can be expressed as a linear amount involving the relevant attributes used in the training examples.

The correlation values reported in Table 4 are not useful for tuning a morphological assessment system. Single traits involving convexity (SC, RC, BC or LL) or angles (RA and RS) do not show a differential pattern when compared to traits involving dimensions. Absolutely linear traits involving smaller anatomical traits such as RL, PW or HL show lower correlation values, while HE, CD and BL show correlations higher than 0·85. However, other traits involving small anatomical traits, such as SC or CW, range between 0·80 and 0·90. In addition, some factors not included in the definition of the traits might also be affecting the assessments. For example, HL might be affected by other head dimensions such as forehead width. Also, muscular mass on the tuber ilii and tuber ischii might affect the

precision of the assessment of RL or RA. Furthermore, LD and LL might be affected by an 'ideal' evaluation of the trait instead of the expected objective assessment. Experts have assessed Asturias de los Valles animals for a long time using a subjective and non-linear methodology. The inertia of the traditional method had to be overcome. In this sense, the methodology proposed in this paper can detect whether experts adjust their outputs to the general specifications of the new linear assessment system.

Some traits showing correlations lower than 0·8, such as RL (correlation 0·68) and RA (correlation 0·64), were correctly defined and understood by classifiers. Supplementary specific training would probably be sufficient to improve the assessment performance in these cases so as to improve the recording of the linear trait assessed.

As a consequence of the aforementioned results, we could not make decisions with respect to the accuracy of the morphological assessment solely through analysis of means and correlations. Therefore, it was necessary to implement a powerful tool to analyse the experts' assessments using, in addition, the objective measurements obtained from the animals.

The average accuracy (error) results obtained by the leaving-one-out method ranged from 0·45 to 1·07, but in general averaged 0·64 for learning systems inducing sets of linear functions (Cubist, SAFE, and M5'), and 0·71 for BETS (Table 4). However, the errors found by the leaving-one-out method are quite similar in all ML systems for each type trait. These errors should thus not be affected by algorithm bias. Traits showing lower ML errors than the average learning system error are found in the skeletal development traits. Some traits with low expected error (RA, LL or HL) show low correlation values. The muscular development traits show higher error values (average errors are 0·76 for Cubist, 0·75 for SAFE, 0·74 for M5', and 0·80 for BETS) than skeletal traits (0·56 for Cubist, 0·57 for SAFE, 0·55 for M5', and 0·64 for BETS) in all ML systems.

The accuracy results show the degree of confidence in ML assessment for each trait. If AI systems learn to assess with a low and reasonable expected error, then we can trust in any of their proposed mechanisms to compute assessments from measurements. In this sense, previous analysis showed that the differences between ML systems and human experts are not significant (López *et al.*, 2000). All the average expected errors are below 1 point. Experts were required to value the trait merits

on a scale from 1 to 9 points. An error of half a point is not a very important event in morphological assessment. Consequently, the performance of the ML systems morphological assessment can be considered consistent. In other words, the experts exhibited quite coherent behaviour from a general point of view.

Another important issue is the number of functions needed to explain the computer assessments. The number of functions that ML algorithms (Cubist, M5', and SAFE) actually produce to derive solutions is a good measure of the linearity of the assessing task. The average numbers of functions (Table 4) for the 1365 learning experiments for each algorithm are 1·57, 1·94 and 2·36 for Cubist, SAFE, and M5' respectively, with similar average error. However, as occurred with errors, in general there are no big differences in the number of functions needed by our ML algorithms.

Most of the skeletal development traits required only one function to be assessed by Cubist and M5'. SAFE also showed a strong trend toward linearity for these traits. Traits involving metric dimensions seem the easiest to assess. Nevertheless, PW, LL and HL needed two functions to be assessed with Cubist and M5'. It is worth noting here that pins width (PW) exhibited the smallest range among all anatomical traits assessed. Expert classifiers can be affected by the age of the assessed animal. They are supposed to be using two different scales, for young and adult animals. The assessment of LL and HL is affected by other considerations apart from the straightness or the length of the head. The classifiers probably have a marked idea of the 'worst' or 'best' for these traits. Classifiers use one scale for concave lines and another for convexity in the LL trait. In the case of HL, the general head dimensions (large and length/large relation) affect the assessment despite the definition of the trait.

Non-linear behaviour was expected for complex muscular development traits (BA, GM, CO). In fact, they needed three or more functions to be assessed by ML systems except CO for Cubist, which explained it as a linear trait. Additionally, the accuracy obtained justifies the hypothesis that these complex muscular development traits could be safely removed from assessment sheets since this animal feature can be computed from a subset of the others.

Further analysis with a greater amount of data can determine a simple rule to assess complex traits such as the compactness of live animals both accurately and objectively. The degree of compactness is a very important issue in carcass assessment (Vallejo *et al.*, 1993). An understanding of the relationship between these two assessments could merit future research.

The traits included in the morphological assessment system for the Asturiana de los Valles breed were defined to show linear behaviour. Consequently, the degree of linearity should be quantified. Given that ML systems are able to produce a collection of linear functions, in general more than one, their number can be used as a realistic measure of the linearity included in expert assessments. If a ML algorithm finally returns only one function to explain the relationship between anatomical measures and type scores, we have strong evidence in favour of the way in which the trait has been measured by human experts. In this sense, single muscular development traits are not so well assessed as skeletal development traits. Traits involving convexity or muscular masses surface require more functions to perform an adequate assessment. These results suggest the possibility of some noise in the design or performance of the assessment system. Therefore, a major effort should be made to improve the computer representation of the traits and achieve a better definition of these.

Implications

The use of AI is a novelty in animal production. The use of AI systems in live animals is troublesome because of the non-uniformity, variability and difficulty of collecting data. Nevertheless, many assessment activities, such as for example conformation recording, could be non-linear. Consequently, human experts' subjective knowledge is invaluable and widely used.

ML algorithms are useful tools for developing conformation assessment systems in beef cattle. The process of knowledge engineering promotes a more useful definition of the traits involved. In addition, AI systems could highlight the sources of inconsistency between the definition and recording of traits. Three sources of inconsistency were detected in the present study: (a) the existence of different interpretations of the definition of the character, increasing the subjectivity of the assessment; (b) the narrow range of variation of some of the anatomical traits assessed; (c) the inclusion of number of complex traits in the assessment system. In this sense, the re-opening of the evaluated Asturiana de los Valles assessment system is recommended.

The detection of these sources of variation will lead to an improved implementation of the assessment systems. An improved definition of the traits involved and an additional effort to train human

experts in the most problematic traits will avoid individual subjectivity in recording data and will improve the accuracy of assessments. In addition, ML systems will allow the study of the relationships among traits. The possibility of obtaining some easy rules to infer the assessment of some complex (and non-linear) traits from some other single trait scores can make the scoring easier and hence the experts' assessments more reliable.

Of course, the implications of AI systems on morphological assessment are further reaching than the present results. The possibility of obtaining a computer assistant tool to teach new classifiers regardless of human errors and to verify and readjust the experts' assessments is easily deduced from the methodology presented in this paper. Moreover, if a representative sample of the assessed animals were scored by the AI system alongside the human assessments, the accuracy of the estimation of environmental factors (particularly the classifier's effect) affecting genetic parameters of morphological traits would increase considerably, thus improving the reliability of the calculated breeding values.

At the moment, the development of the application of the AI technology to conformation recording is affected by the difficulty of acquiring digital images of the animal (López *et al.*, 2000). Additionally, the preparation of the sets of examples to train ML algorithms is a laborious task. Therefore, there is a need to design hardware suitable for use in farm conditions to acquire digital images of the animal so as to be able to carry out the necessary zoometry in a safer and painless way with reduced costs or logistics overheads, especially for breeds lacking the docile behaviour characteristic of the Asturiana de los Valles animals.

Acknowledgements

References

Associazione Nazionale Allevatori Bovini Italiani Carne. 1997. *La razza Chianina.* Tipolitografía Grifo, Perugia.

Bahamonde, A., de la Cal, E. A., Ranilla, J. and Alonso, J. 1997. *Self-organizing symbolic learned rules.* Lecture Notes in Computer Science, LNCS no. 1240, pp. 536-545. Springer-Verlag, Berlin.

Berg, R. and Butterfield, R. 1979. *Nuevos conceptos sobre desarrollo de ganado vacuno.* Editorial Acribia, Zaragoza, Spain.

Brotherstone, S. 1994. Genetic and phenotypic correlation between linear type traits and production traits in Holstein-Friesian dairy cattle. *Animal Production* **59:** 183-187.

Brotherstone, S. and Hill, W. G. 1991. Dairy herd life in relation to linear type traits and production. I. Phenotypic and genetic analyses in pedigree type classified herds. *Animal Production* **53:** 279-287.

Cover, T. M. and Hart, P. E. 1967. Nearest neighbour pattern classification. *IEEE Transactions on Information Theory* **13:** 21-27.

de la Fuente, L. F., Fernández, G., and San Primitivo, F. 1996. A linear evaluation system for udder traits of dairy ewes. *Livestock Production Science* **45:** 171-178.

del Coz, J. J. and Bahamonde, A. 1999. Mapas autoorganizados con atributos discretos. *Proceedings of the VIII Conferencia de la Asociación Española para la Inteligencia Artificial, Murcia, Spain, vol. I,* pp. 117-124.

del Coz, J. J., Luaces, O., Quevedo, J. R., Alonso, J. and Bahamonde, A. 1999. *Self-organizing cases to find paradigms.* Lecture Notes in Computer Sciences, LNCS no. 1606, pp. 527-536, Springer-Verlag, Berlín.

Fernández, G., Alvarez, P., San Primitivo, F. and de la Fuente, L. F. 1995. Factors affecting variation of udder traits of dairy ewes. *Journal of Dairy Science* **78:** 842-849.

Goyache, F., Villa, A., Baro, J. A. and Alonso, L. 1999. Aplicación de un sistema de calificación morfológica continua en la raza Asturiana de los Valles. *Federación Española de Asociaciones de Ganado Selecto* (*FEAGAS*) **16:** 8-17.

International Committee on Animal Recording. 1995. *Recording guidelines: appendices to the international agreement of recording practices. Section 5: Conformation recording.* Rome, Italy/RVN, Arnhem, the Netherlands.

Linko, S. 1998. Expert systems — what can they do for the food industry? *Trends in Food Science and Technology* **9:** 3-12.

López, S., Goyache, F., Quevedo, J. R., Alonso, J., Ranilla, J., Luaces, O., Bahamonde, A. and del Coz, J. J. 2000. Un sistema inteligente para calificar morfológicamente bovinos de la raza Asturiana de los Valles. *Revista Iberoamericana de la Inteligencia Artificial* **10:** 5-17.

Luaces, O., del Coz, J. J., Quevedo, J. R., Alonso, J., Ranilla, J. and Bahamonde, A. 1999. *Autonomous clustering for machine learning.* Lecture Notes in Computer Sciences, LNCS no. 1606, pp. 497-506. Springer-Verlag, Berlin.

Michalski, R. S., Bratko, I., and Kubat, M. 1998. *Machine learning and data mining, methods and applications.* John Wiley and Sons Ltd, Chichester.

Nilsson, N. J. 1998. *Artificial intelligence: a new synthesis.* Morgan Kaufmann, San Francisco, CA.

Pietersma, D., Lacroix, R., Lefebvre, D., Block, E. and Wade, K. M. 1999. Example-based knowledge acquisition for automated interpretation of milk recording data. *Proceedings of the second European conference of the European Federation for Information Technology in Agriculture, Food and Environment, Bonn, Germany,* pp. 669-677.

Quevedo, J. R. and Bahamonde, A. 1999. Aprendizaje de funciones usando inducción sobre clasificaciones discretas. *Proceedings of the VIII conferencia de la Asociación Española para la Inteligencia Artificial, Murcia, Spain, vol. I,* pp. 64-71.

Quinlan, J. R. 1993a. *C4·5: programs for machine learning.* Morgan Kaufmann, San Mateo, CA.

Quinlan, J. R. 1993b. Combining instance-based and model-based learning. *Proceedings of the tenth international machine learning conference.* Morgan Kaufmann, Amherst, MA.

Quinlan, J. R. 2000. Cubist release 1·08, http://www.rulequest.com/cubist-info.html

Rich, E., and Knigh, K. 1991. *Artificial intelligence, second edition.* McGraw-Hill, New York, NY.

Shi, M. J., Laloè, D., Menissier, F. and Renand, G. 1993. Estimation of genetic parameters of preweaning performance in the French Limousine cattle breed. *Genetics, Selection, Evolution* **25:** 177-189.

Short, T. H. and Lawlor, T. J. 1992. Genetic parameters of conformation traits, milk yield and life in Holsteins. *Journal of Dairy Science* **71:** 1987-1998.

Vallejo, M., Gutiérrez, J. P., Cima, M., Cañón, J., Alonso, L., Revuelta, J. R. and Goyache, F. 1993. Características de las canales de las razas bovinas asturianas. III. Valoración cuantitativa y predicción tisular de canales en la raza Asturiana de los Valles. *Archivos de Zootecnia* **42:** 29-40.

Vukasinovic, N., Moll, J. and Künzi, N. 1997. Factor analysis for evaluating relationships between herd's life and type traits in Swiss Brown cattle. *Livestock Production Science* **49:** 227-234.

Wang, Y. and Witten, I. H. 1997. Inducing model trees for predicting continuous classes. *Proceedings of the European conference on machine learning.* University of Economics, Prague.

Yang, X. Z., Lacroix, R. and Wade, K. M. 1999. Neural detection of mastitis from dairy herd improvement records. *Transactions of the American Society of Agricultural Engineers* **42:** 1063-1071.