

The usefulness of Artificial Intelligence techniques to assess subjective quality of products in the food industry

F. Goyache†, A. Bahamonde‡, J. Alonso‡, S. López‡, J.J. del Coz‡, J.R. Quevedo‡, J. Ranilla‡, O. Luaces‡, I. Alvarez†, L.J. Royo† and J. Diez†

† *SERIDA-CENSYRA-Somió, C/ Camino de los Claveles s/n,
E-33208 Gijón (Asturias), Spain. Tel: +34-98 5195303. E-mail: felixgg@princast.es*
‡ *Centro de Inteligencia Artificial. Universidad de Oviedo in Gijón, Campus de Viesques,
E-33271 Gijón (Asturias), Spain. Tel: +34-98 5182032.
E-mail: { antonio, jalonso, secun, juanjo, quevedo, ranilla, oluaces }@aic.uniovi.es*

Abstract

In this paper we advocate the application of Artificial Intelligence techniques to quality assessment of food products. Machine Learning algorithms can help us to: a) extract operative human knowledge from a set of examples; b) conclude interpretable rules for classifying samples regardless of the non-linearity of the human behaviour or process; and c) help us to ascertain the degree of influence of each objective attribute of the assessed food on the final decision of an expert. We illustrate these topics with an example of how it is possible to clone the behaviour of bovine carcass classifiers, leading to possible further industrial applications.

Postal address: F. Goyache, Sección de Mejora Genética
CENSYRA-SERIDA-Somió
C/ Camino de los Claveles s/n, E-33208, Gijón (Asturias), Spain.

Phone: +34 985 19 53 03
Fax: +34 985 19 53 10
E-mail: felixgg@princast.es

The usefulness of Artificial Intelligence techniques to assess subjective quality of products in the food industry

F. Goyache†, A. Bahamonde‡, J. Alonso‡, S. López‡, J.J. del Coz‡, J.R. Quevedo‡, J. Ranilla‡, O. Luaces‡, I. Alvarez†, L.J. Royo† and J. Diez†

† *SERIDA-CENSYRA-Somió, C/ Camino de los Claveles s/n,*

E-33208 Gijón (Asturias), Spain. Tel: +34-98 5195303. E-mail: felixgg@princast.es

‡ *Centro de Inteligencia Artificial. Universidad de Oviedo in Gijón, Campus de Viesques,*

E-33271 Gijón (Asturias), Spain. Tel: +34-98 5182032.

E-mail: { antonio, jalonso, secun, juanjo, quevedo, ranilla, oluaces }@aic.uniovi.es

Abstract

In this paper we advocate the application of Artificial Intelligence techniques to quality assessment of food products. Machine Learning algorithms can help us to: a) extract operative human knowledge from a set of examples; b) conclude interpretable rules for classifying samples regardless of the non-linearity of the human behaviour or process; and c) help us to ascertain the degree of influence of each objective attribute of the assessed food on the final decision of an expert. We illustrate these topics with an example of how it is possible to clone the behaviour of bovine carcass classifiers, leading to possible further industrial applications.

Introduction

The food industry needs to supply markets with safe, high-quality products to satisfy consumer demands. The design of standard, reliable procedures to control the quality of products is therefore a major objective. However, the food industry usually works with non-uniform, variable raw materials, and processes that exhibit highly non-linear behaviour.

In many cases, the cumulated experience of market operators (sellers and buyers) has reached a rough consensus on the way of grading the quality of a given product. This consensus is supported by the agreement on general non-systematic rules that are performed by experts. The performance of the experts is, in turn, based in their own experience. Since it is useful to market operators and acceptable to consumers, modern markets need to generalise quality-grading procedures obtained from subjective human behaviour. However, training of classifiers is difficult and expensive. They need to cumulate a large experience before to carry out commercial classifications and they must be retrained periodically to avoid substantial individual differences in assessments. Notwithstanding the complexity of training procedures, the repeatability of grading tends to be low, thus affecting the market's confidence. In consequence, the implementation of techniques that could perform cheaper, accurate and repeatable assessments according with the well accepted human expertise is an interesting task for food industry.

Linko [1] advocates the introduction of expert systems in the food industry as a way to obtain more reliable automatic control of food processes. Expert systems are expected to incorporate human expert knowledge into quality control process. However, this is only possible when expertise is clearly expressed in well-established rules, able to be performed by a computer routine. The bottleneck to build expert systems is that frequently, human expert knowledge cannot be summarised systematically. Experts do not know how (or they do not want) to explain the way they make decisions affecting quality grading. Frequently, the expert's behaviour is expressed in general rules that are difficult to apply in a routine way. In this situation, the food industry is affected since the implementation of expert systems to automation of processes is not easy.

The beef industry is an evident example of this situation. The European Union set up a standard bovine carcass conformation assessment system, known as the SEUROP system, to be applied in the whole EU territory (CEE 390/81, CEE 1208/81, CEE 2930/81 and CEE 1026/91 rules). This grading system is expected to be useful for

bovines from 300 Kg of live weight regardless of sex, age or management conditions of the animal. Of course, the SEUROP system is described in a very broad way. The conformation class description ranges from 'exceptional' to 'poor' muscular development and from 'extremely' convex to 'very' concave profiles. Under these conditions, it is not surprising that attempts to automatically classify bovine carcasses using data such as carcass weight and linear measurements as objective input could not be considered as successful [2]

On the other hand, the beef industry would be interested in grading the final produce (meat) taking into account sensorial quality. However, we do not know how to explain the decisions of a sensory panel. Sensorial quality grading is dependent on a not well defined group of properties resulting from natural structure of the food elements, their mutual arrangement, interaction and way in which they are perceived by the human senses [3]. Sensorial perception can integrate a large number of single sensations at the same time. The complexity of the sensorial performance of food products can be simulated during instrumental measurements, but a complete reproduction of sensory testing is impossible. There are a large number of instrumental methods to estimate sensorial quality. Frequently, there are various laboratory tests to characterise a quality variable. However, correlations between instrumental measurements and sensory panel decisions are usually low [4, 5, 6, 7] and we cannot easily know whether a laboratory determination is the appropriate one to evaluate sensorial quality.

The situation affecting quality assessment in food produce that we have tried to illustrate suggests, as an interesting topic, the possibility of implementing different tools based on Artificial Intelligence (AI in the following) techniques in the food industry. AI can be seen as that part of Computer Sciences that tries to simulate processes that would be described as intelligent behaviour in humans. Thus, Machine Learning (ML in the following) is one of the central topics of AI, since a feature usually attached to intelligence is the ability to learn from the environment. From a general point of view, ML algorithms synthesise knowledge from an unorganised source in such a way that their outputs are in fact computer programs (i.e. expert systems) able to accomplish useful tasks. To mention some examples, ML technologies have been used to design and improve subjective assessment systems in animal production [8] and to estimate sensory panel evaluation of food more accurately than classical statistical treatment [9].

It is interesting to highlight the fact that AI can contribute to the food industry in a broader way than the sole building of expert systems to automate and monitor the control of food quality. We would like to highlight how ML methodologies can help researchers and technicians to ascertain what the behaviour of humans is in order to decide the subjective quality of food products and to decide what objective attributes of food are determinant to obtain a classification according to market and consumer demands. In other words, we will try to emphasise the way that ML algorithms can help us to: a) extract operative human knowledge from an easily available source, i.e. a set of examples; b) conclude interpretable rules for classifying samples regardless of the non-linearity of the human behaviour we are studying; and c) ascertain the degree of influence (relevancy) of each objective attribute of the assessed food on the final decision of an expert. We will illustrate the behaviour of ML algorithms with an example of how it is possible to successfully clone the behaviour of bovine carcass classifiers, determine the degree of linearity of their decisions and finally, determine the smallest amount of objective data needed to make an accurate carcass classification.

Artificial Intelligence techniques

AI tries to clone human behaviour. To do so, AI deals with knowledge like conventional computer programs deal with data. To be used as input for learning functions, knowledge must be "represented". In the AI context, representation stands for translating something as abstract as knowledge into a formal system of symbols that can be stored on a computer. Thus, knowledge representation is a central subject in AI [10, 11]. Knowledge-based systems are hence programs that manage these symbol systems to solve practical problems. In the case of ML, we must handle knowledge both as an input (although unstructured) and as the output in the form of a computable prediction function.. Thus, ML algorithms are conceptually the same as linear regressions; the differences are however quite important. Prediction functions obtained from ML algorithms can be more sophisticated than simple linear formulas; they are close approaches to more general (not linear) functions. But there is another important difference; ML algorithms can handle categorical labels as both attribute (independent variables) and class values. Therefore the application field of ML algorithms and their

output possibilities are broader than with simple linear least squares tools. A detailed explanation of ML principles and methods can be found in textbooks such as Nilsson [10], Rich and Knight [11], Quinlan [12] or Michalski *et al.* [13].

The most common input 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. 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).

To build the training sets to learn how to assess the quality of food products, we must first represent the relevant attributes involved in perceiving product conditions. These relevant attributes are the “objective” data we use to feed AI algorithms. This is a difficult task, since the objectivity and faithfulness of this representation will determine the usefulness of the results found by AI algorithms. Sometimes we do not have sufficient sources of information to represent our classes correctly; in addition, we often do not know whether some available information is relevant or “noise”. For instance, sometimes there is a bundle of geometrical shapes that must be understood, as happens when we want to assess bovine carcasses. Hence, each geometrical concept must be represented by means of a set of mathematical formulas to quantify the shapes. But this is only the first half of the expertise that we must handle in order to learn; we must attach the score that our experts consider the right one to be assigned to each item. In general, we must attach to the representation of each event the kind of behaviour that we would like to cause.

After we build the training sets we should select the type of ML algorithm we think are the most interesting to obtain the best performance. There are several classes of ML algorithms. Different kinds of algorithms can accurately solve a given problem. We can decide on the use of one algorithm taking into account their kind of outputs.

The degree of explicitness of the output function is frequently taken as a classification dimension for ML algorithms. The most interpretable representations are given by those systems (AI algorithms producing decision trees or sets of rules) that conclude a collection of regression rules, i.e. linear functions endowed with the conditions to be applied to estimate or predict the class. ML algorithms that return complete regression rules are Quinlan’s M5 [14, 15] or its *rational reconstruction* M5’ [16], which conclude decision trees with linear functions on their leaves. The commercial algorithm Cubist [17] and our SAFE [18] provide a list of regression rules (see Table 2). At the other end of the spectrum are to be found the opaque computational devices that become artificial neural networks when they are trained. The nearest-neighbour algorithms [19, 20, 21] are probably to be found in between.

At the same time, for systems producing regression rules, 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 symbols, the structure of a regression rule is the following:

Rule R :: class = $f_R(x_i; i \in I_R)$ if <conditions>

where conditions is a conjunction of requirements about the numeric values of a subset of the attributes, and f_R is a linear function involving a subset I_R of continuous attributes; notice that this includes constant functions. Usually, **conditions** will look like

att1 \in [a, b] & att3 \in [c, d]

In the exact agreement method, all of the conditions must be true to make use of its associated prediction function or value. So, in the preceding example of Rule **R**, an example \mathbf{x} fulfilling this rule must have values of attribute **att1** (usually represented by x_1) between **a** and **b**, and **att3** (x_3) between **c** and **d**; in this case the predicted class of \mathbf{x} will be provided by applying f_R to the corresponding \mathbf{x} attribute values. On the other hand, if the membership requirements are relaxed, we have fuzzy evaluators. In this case, the rule selected to evaluate a new case \mathbf{x} need not necessarily fulfil all its conditions completely, but rather its application conditions should be more similar to \mathbf{x} values than any other rule. The predicted class is then computed interpolating the values provided by the nearest regression rules. Obviously, for fuzzy evaluators the metric

used to measure the similarity between two cases is very important. This evaluation mechanism is typical of systems that remember prototypes as the nearest neighbour algorithms.

There are some ML algorithms, such as Artificial Neural Networks (ANNs) and genetic algorithms, that are becoming popular far removed from the AI environment, and which can be useful to obtain accurate classification. ANNs are widely used when explicitness is not a requirement. Here the training goal is a matrix of weightings attached to the arcs of a directed graph biologically inspired by natural neural networks. Since these graphs compute numerical functions, the matrix is iteratively adapted to produce similar returns to training input-output pairs; i.e. training examples. ANNs could be used to evaluate the difficulty of learning how to assess food quality, but ANNs do not return explicit functions. In addition, ANNs need the previous definition of the layout and other parameters, and so accuracy estimations could be affected by deficiencies in the initialisation arrangement instead of the *learnability* of training sets. The problem is that there are no automatic procedures to guess a reasonable ANN layout. However, it must be clear that the input of ANNs is the same as has to prepare to use other explicit learning methods; so a knowledge representation has to be designed and implemented in any case.

On the other hand, genetic algorithms are not strictly a ML method or algorithm. In fact, they are search methods that in turn have been used thoroughly to learn in a very general sense, but have been implemented in an *ad hoc* fashion for each problem. Once more, our aim was not to present comparison studies, nor to introduce specific algorithms devised to solve assessing, as we could do with the assistance of genetic algorithms. In other words, the goal of this paper is not suitable for using ANNs or genetic algorithms.

In this paper we try to highlight the interest in seeing what was learned in the quality assessment environment. The explicit expressions of the rules are interesting in themselves.

Machine Learning outputs

In this section we will explain the kind of outputs the user can expect when using ML algorithms. Let us stress that in addition to the prediction functionality, the knowledge synthesised by these algorithms can be useful due to their explanatory performance. The most interesting characteristics we can observe are:

- a) the possibility of induction of non-linear assessing criteria
- b) the possibility of obtaining explicit explanations of what is learned
- c) the possibility of determining the most important attributes affecting assessment criteria; in other words, the possibility of knowing the relevancy of the attributes used to represent a problem

Despite the fact that we are usually forced to consider food processes under a linear performance, most food product assessing criteria can follow a non-linear behaviour. However, the determination of the linearity of a given behaviour is a subtle matter. A first step in tackling this question is to compute the correlation coefficient. But unfortunately, this is not a wise decision, since it is not easy to define a threshold to discriminate (by its correlation coefficient) which variables in fact have a linear relationship or not. The geometrical reason is that a set of points defining 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 has to explain the behaviour of numerical relationships. Figure 1 depicts such a situation, where graph A shows a clear linear behaviour while graph B does not. However, the correlation coefficient of the graph A is $r = 0.81$, while the correlation coefficient of graph B is $r = 0.96$. In this situation, regression is not a good predictor of linearity of the process under study. Of course, we could obtain lower correlations if we decreased the angle formed by the linear segments.

The usual approach of ML algorithms can be seen as a clustering that tries to aggregate points (in fact training examples) of reasonably uniform behaviour. Therefore, the output of ML algorithms adapt better than simple linear regression to functions where the formulae to be applied depend on the circumstances described by any

of the attribute values. This characteristic of AI outputs can be very interesting in the case of the evaluation of assessing criteria of food products.

As mentioned above, the degree of explicitness of ML outputs is a way to classify these algorithms. When we try to assess the quality of a kind of item, the reward for obtaining an explicit representation of a prediction function is that we can then endow predictions with some explanations of why we are postulating the scores. We can simply provide the quality value in addition to the part of knowledge used to compute that number or label. These explanations may be meaningful or not, depending on the conceptual clarity of the knowledge and the similarity with the kind of reasoning used by human experts when they are carrying out the same task.

In fact, explanations are not always useful. On production lines, for instance, automatic quality control devices do not need to explain their decisions when they are sufficiently tested in laboratory conditions. However, if we would like to obtain control advice about the way to improve production, we need the control system to be able to elaborate readable explanations of the cause of their assessment decisions.

Another utility of the explanatory capabilities of what is learned is that we can rewrite the classification rules as procedure manuals to train future human experts. This is the case when the computerised descriptions of the item to be classified is very difficult or expensive to obtain, and it is only possible to collect an experimental sample in order to induce a compact and sound group of explicit rules.

Frequently, we do not know what the most important attributes are to solve problems concerning knowledge. In fact, ML techniques are often used when factors affecting a process are not well known. This is the case, for instance, of sensory panel decisions. We can describe quality-grading decisions with a broad list of attributes. Of course, not all the attributes have the same weighting to grade a set of examples. Sometimes their contribution is only redundant or negligible information. It is clear that if it were possible to ascertain the most discriminate attributes affecting a problem, their solution would be easier.

So, when we are trying to induce knowledge, the study of the relevancy of the attributes that describe the examples is a core issue. Many ML algorithms include some type of mechanism to select the major attributes affecting the studied knowledge. In the AI arena there are two main streams. The first one looks for weightings of the attributes values to be used when we measure the similarity of observed cases in the past to solve present situations [22, 23]; usually, the assigned weightings range between 0 and 1. On the other hand, there are the proposals that search for a subset of relevant or not redundant attributes to be taken into account by any ML algorithm [24]; these methodologies select the more relevant attributes, removing the rest. In both cases, the underlying idea is to be able to estimate the prediction quality of the attribute or attribute values in a ML environment.

However, the frontiers between both methodologies are not pure. In fact, there are algorithms that first use a weighting assignment to select the more relevant attributes in a second step [25].

An example: automatic assessment of bovine carcasses

We will try to illustrate the topics highlighted in the sections above with a real problem: bovine carcass conformation assessment. We think this example may show the usefulness of Artificial Intelligence with respect to more traditional methods because:

- a) Bovine carcass grading following the SEUROP system is compulsory in the whole territory of the European Union, being governed by strict regulations: CEE 390/81, CEE 1208/81, CEE 2930/81 and CEE 1026/91.
- b) The SEUROP system is expressed in the form of very broad rules that are not easy to perform in a systematic way
- c) There is a lack of agreement on how to represent the carcass conformation of bovines objectively. There are a large number of possible carcass measurements [26, 27]. Although the European Association of

Animal Production officially adopted the methodology proposed by De Boer *et al.* [28], there is no clear agreement on its use [29].

- d) The use of non-intelligent systems to assess bovine carcass conformation is not well developed and human expertise is widely used [30, 31].

The aim of the present application is to ascertain the major attributes affecting carcass-grading decisions to enable the future implementation of a computer system to assess bovine carcass conformation capable of being used in an industrial environment with little interaction with an operator.

The proposed problem needed a major representation effort. To reach our objective, we benefited from the valuable collaboration of three expert classifiers of EASA, the control organisation of the Quality Beef Program “Carne de Asturias”. After thoroughly discussing with our experts the way they implement the SEUROP classification, we concluded that traditional measurements cannot accurately represent carcass conformation. Expert classifiers considered the influence of some attributes as very important: muscular profiles, muscular development of different carcass sections or ratios between different measurements (for example, between back length and rear or front back width). We concluded that a set of 28 attributes could numerically represent the features mentioned in SEUROP specifications. These 28 attributes, including hot carcass weight and the blockiness index, calculated as the ratio between carcass weight and carcass length [32], are listed and described in Table 1

As far as it is impossible to perform most of these measurements in the slaughterhouse, we developed an application, based on our previous experience with live beef cattle conformation [8], to calculate said measurements from digital images obtained by our experts of the assessed carcasses [33]. Each carcass was photographed in three different positions: (a) lateral, (b) medial and (c) dorsal (see Figure 2). We included a metric reference in each photo so as to be able of measure lengths from the pictures. These images were then processed by marking 21 key points and 5 curve arcs (Figure 2) to calculate the 28 attribute values. Single anatomical traits were easily calculated by means of lengths (i.e. carcass length = distance between I2 and I7 in Figure 2), ratios and volumes (to estimate muscular development) combining some single anatomical traits. The precision of the attained digital measurements was successfully tested by comparing them with standard measurements taken on the half-carcass according to the methodology described by De Boer *et al.* [28]. We did not observe any bias due to the subjectivity of the operator who marked the points on the images [33].

The representation of lengths and volumes is straightforward, but profiles representation needs some explanation. Given a profile such as EC1 in Figure 2 (a), we consider the curve arc that borders it (d) as a variable real function \mathbf{f} with respect to some axes with their origin at the left hand side of the curve arc (e). We can then compute the curvature at each point $(\mathbf{x}, \mathbf{y} = \mathbf{f}(\mathbf{x}))$ by means of the formula

$$\text{curvature}(\mathbf{x}) = \frac{\mathbf{f}''(\mathbf{x})}{(1 + (\mathbf{f}'(\mathbf{x}))^2)^{3/2}}$$

Since we do not actually have the explicit formula of the curve representing a profile, we must approximate the derivatives using the values of \mathbf{f} in the environment of each point. So we divide the arc by means of a sequence of points $\{\mathbf{x}_i\}$ in $[0, \mathbf{a}]$ that divide the interval in a given number (the same in all cases) of equal length subintervals. Then $\mathbf{f}'(\mathbf{x}_i)$ and $\mathbf{f}''(\mathbf{x}_i)$, the first and second derivative, are approached using

$$\mathbf{f}'(\mathbf{x}_i) = \frac{\mathbf{f}(\mathbf{x}_{i+1}) - \mathbf{f}(\mathbf{x}_{i-1})}{2} \quad \text{and} \quad \mathbf{f}''(\mathbf{x}_i) = \frac{\mathbf{f}'(\mathbf{x}_{i+1}) - \mathbf{f}'(\mathbf{x}_{i-1})}{2}$$

Finally, to summarize the convexity of the whole arc in the interval $[0, \mathbf{a}]$ (see Figure 2 (e)), we compute the average of the $\text{curvature}(\mathbf{x}_i)$ for all $\{\mathbf{x}_i\}$.

After we considered the way to computationally represent bovine carcass conformation to be satisfactorily fitted, we built a training set to feed ML algorithms. Our experts photographed 104 bovine carcasses, and then, following the SEUROP system, graded carcasses individually. To improve the accuracy of classification

it was permitted to add the signs + or - to each conformation grade [29, 34]. Each conformation grade was scored numerically from 1 (P) to 6 (S). The signs + and - were computed as +0.25 or -0.25 conformation points, respectively, on the numerical conformation score. Most carcasses were graded by each classifier. In fact, the first one classified 84 carcasses, the second 79 and the third 80. All this sums 243 classification events, which were added to the training set as independent training examples. The distribution of the conformation classes of the training set was: 9, S; 128, E; 64, U; 32, R; 8, O, and 2, P. A sizeable proportion of examples were from females (29%) and from light carcasses, weighting less than 220 Kg (28%). This training set is considered a representative sample of the real distribution of sexes, conformation classes and weights existing in the beef market in Asturias.

As was highlighted before, different kind of ML algorithms can perform differently when faced with of a particular problem. To illustrate the topics we are presenting in this report, we used the following ML algorithms:

- Cubist [17] and M5' [13; 16]. These are ML algorithms that induce regression rules that make up a crisp evaluation mechanism.
- SAFE (System to Acquire Functions from Examples) [18]. This ML system returns a set of regression rules (see Table 2) that define linear functions to be applied when the attached conditions are fulfilled. When more than one rule can be applied, the order in which rules are listed defines the priority. However, when the case does not fulfil any of the rule conditions, the predicted score for that case will be computed by interpolating (inversely to their distance to the case) the values provided by the nearest rules; i.e. SAFE uses fuzzy evaluation. In the example shown in Table 2, 230 examples from the training set are covered by the 9 rules learned by Safe; the other 13 examples have to be evaluated by interpolation.
- BETS (Best Examples in Training Sets) [21, 35, 36] is a ML system of another family. From a conceptual point of view, BETS induces regression rules where linear functions are always constant, and conditions are defined by intervals with only one point. In straightforward terms, BETS' regression rules are simple examples selected from the training set where the attribute values considered irrelevant are skipped.
- LC3 (Learning to Calculate Continuous Classes) [37] selects, as does BETS, a subset of training examples. It is a continuous version of Domingos' RISE [38], with the addition of a heuristic to measure the prediction quality of regression rules based on the impurity level [39].

To estimate the accuracy of the performance of the algorithms, we considered the average conformation score of the three experts as the class we would like to learn. We calculated the differences between the conformation score calculated by our algorithms (Table 3) through a *cross validation* and we compared the average of the calculated differences with the differences between the score of each classifier and the average of the three classifiers. Notice that the average absolute differences between the scores of each expert and the average conformation score was 0.41 for the first, 0.39 for the second and 0.41 for the third. Additionally, we compared the ML algorithms accuracy with differences between the target class and the conformation score calculated by means of classical linear regression.

With these assumptions we expect the ML algorithms performance to be penalised because individual classifier scores are used to estimate the target class. However, with the *cross validation* system we divided the training set into 10 folders and we successively used each of these folders as a test set while the other 9 were used for training; i.e. for each example of the test folder we applied the prediction function obtained by the ML algorithm from the other 9 folders, and we then computed the absolute difference with respect to the class of the example. We repeated this experiment 5 times, finally returning the average of the differences thus computed as 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. The same methodology was employed for the classical linear regression. Table 3 reports the estimations of absolute errors for all the systems used. Table 3

also shows the average number of functions or prototypes the algorithms need to learn in order to grade bovine carcasses. The number of linear functions used by algorithms producing a set of rules (SAFE, Cubist and M5) provides an overall estimation of the linearity of the assessing behaviour. Table 2 shows the single formula obtained by simple linear regression and the 9 formulas calculated by SAFE [18] to assess the SEUROP classification for bovine carcass, stressing a high degree of non-linear behaviour. The training examples set has 243 elements and is analysed in the rest of the paper. The number of prototypes used by algorithms selecting examples from the training set (BETS and LC3) provides an overall estimation of the difficulty of the learning behaviour.

The learning experiments were carried out with different training sets after we applied techniques to estimate the relevancy of attributes. The aim, as explained above, was to extract a small subset of relevant attributes from which to compute the SEUROP conformation score without penalising the accuracy of the assessment. Thus, we ordered the attributes according to their prediction power by means of the tools developed from SAFE and BETS. The experiments reported in Table 3 show three possible ways of describing a bovine carcass: 28 attributes (the whole set), 15 and 4 attributes (excluding the least relevant). The most relevant attributes were, in this order: belly depth, compactness index, and round and topside profiles.

Since conformation score ranges within a conformation grade between +0.25 and -0.25, average differences higher than 0.5 may be considered unacceptable. The learning target would be 0.40, which is the average mean of the classifiers' errors.

Despite the good behaviour of linear regression, Table 3 shows the superiority of the AI techniques with respect to classical methodologies in obtaining the proposed goals. Using 15 or more attributes, the average error obtained using linear regression is acceptable and comparable with the ML algorithms error, with the exception of M5 behaviour using 15 attributes, which return an accuracy of 0.40. Here we must stress the important role played by knowledge representation techniques used to translate into numbers what our experts and the SEUROP manuals explain. We obtained scores for very important variables, such as profiles, which have improved the estimation of conformation scoring using classical statistical tools. We may say that carcass conformation assessment, despite its clearly non-linear behaviour, is learnable using linear regression in a theoretical setting. However, the use or estimation of 15 attributes is impractical in an industrial environment.

In this sense, the superiority of AI techniques is clear with respect to ascertaining the major factors affecting the studied process. If we are concerned with the usability of the automatic assessment, we must reduce the number of attributes used to represent a bovine carcass. If we consider the attributes showing a significant correlation coefficient with the carcass score as relevant in order to obtain a good carcass-grading prediction, we would have to consider the use of 11 different factors, which is still impractical. The results reported in Table 3 show that 4 attributes can be sufficient to predict SEUROP scores with a deviation of 0.44 if we use LC3 rules, or just 0.45 for BETS or Cubist. Notice that even with only 4 attributes, the expected deviation is quite similar to that of the experts (0.41, 0.39 and 0.41 respectively), although the opinion of each expert is 1/3 of the consensus score that we used to compute their own deviation. The better adjustment of ML algorithms to non-linear behaviours (Table 2) means it is possible to maintain a good prediction performance when the number of available data decreases. On the other hand, let us stress that with only 4 attributes, the abilities of linear regression methods are clearly insufficient; the deviation of 0.62 is, as mentioned above, not acceptable.

An important consequence of the use of the 4 selected attributes (belly depth, compactness index, and round and topside profiles) is that we would only need to obtain 2 photos and mark 8 points to achieve useful carcass grading. In fact, the belly depth (distance between I4 to I5) requires 2 points; the compactness index only needs to compute carcass length (distance between I2 to I7) since carcass weight can be provided by a mechanical device in slaughterhouse; profiles of round and topside can be computed if we mark the start and end points of the respective arcs. The implementation of a device to obtain scores of these 4 attributes from digital images in standard conditions without the intervention of human operators seems to be an attainable goal, but this is far from the aim of this paper. Let us mention here that although the round profile was originally considered in the lateral photo (see Figure 2), it is trivial for us to use the medial photo, given that

the profile in this case is given by a function of the form $g(x) = f(a-x)$ (see Figure 2), where f is the function of the arc in the exterior, and thus g and f have the same curvature in the interval $[0, a]$.

Conclusions and future research trends

There is general consensus in the food industry regarding the advantages of using standard and automatic procedures to obtain more reliable automatic control of food processes. These advantages have been summarised by Linko [1]: 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 hours a day; e) they can operate in risky situations; f) they can act quickly on the basis of huge databanks and knowledge banks; and g) they can use natural language, and do not require complex mathematical expressions.

The aim of this paper is not to discuss these advantages, but rather to highlight the possibilities of using a new tool in food research and technology: namely AI techniques and ML algorithms. The main advantages of using AI techniques are: a) they are adapted to working in a non-linear behaviour, as generally occurs in food processes, b) they can explicitly explain what is learned; and c) they can be used to ascertain the major attributes affecting process performance.

All these characteristics of AI techniques allow sound assessments to be obtained, comparable to accepted human performance, when classical statistical tools are not handy. At the same time, the rules obtained from learning procedures can be rewritten as procedure manuals to train future human experts. This is the case when the computerised descriptions of the item to be classified is very difficult or expensive to obtain, and it is only possible to collect an experimental sample in order to induce a compact, sound group of explicit rules. But probably the most interesting use of AI is to ascertain what the most important attributes are in order to solve problems concerning knowledge when we do not clearly know what the core of the process is.

The example presented in this paper illustrates these major characteristics of ML techniques. Carcass grading is usually considered as an activity based on human expertise. Non-linearity of grading makes it difficult to clone the well-accepted human behaviour using traditional tools. However, AI exhibits acceptable performance in spite of the reduction of the number of attributes available to obtain an accurate assessment. The number of attributes needed to assess carcass conformation enables further development of methodologies useful in industrial behaviour to obtain accurate results. Following the assumptions of our methodology, only 8 points and 2 photos of each carcass are needed to guarantee a deviation from the consensus conformation score similar to the deviation of our experts. Even though a human operator would be needed to process the carcass' digital images, no more than 10 seconds were spent to carry out the process. To reach a similar performance using simple linear regression, a costly additional effort would be needed.

Of course, we do not generalise the results presented in this paper with respect to the example of bovine carcass conformation grading. These results were obtained in a particular market under specific industry conditions. Our methodology should be tested in a wide number of different environments, gathering more and more different human expertise.

The use of AI techniques exemplified in this paper includes learning computable ways to attach assessments (numerical scores) of the quality of food products. Typically, these scores have a straightforward numeric, semantic, discrete or continuous value; i.e. an integer or real number. However, the attributes that define the product to be assessed may sometimes have values such as colours or tastes or any other linguistic variable, whose numeric coding may be arbitrary.

In this sense, one of the most interesting fields for further application of AI techniques in the food industry will be the study and cloning of the behaviour of sensory panels used to decide the consumer quality of food. Sensorial quality depends on many factors such as: chemical composition, structure, physical properties, processing methods, shape and many others, with mutual relationships and intensity designating this property as complex and thereby difficult to univocally define. AI behaviour seems to be able to clone the complexity of human decisions. With respect to the beef sector, there are a large number of instrumental methods to

estimate sensorial quality of meat [40, 41, 42]. They are less time-consuming, cheaper, more repeatable, and free from variability caused by psychological, physiological and environmental factors influencing human responses. However, we do not know what the most important methods are to accurately represent human sensorial decisions [43, 44]. Over the last decade, a large number of research groups have created important databases containing many laboratory (physical or chemical) determinations of meat characteristics affecting a large number of cattle breeds, production conditions and markets requirements. These determinations have usually been conducted simultaneously with sensory panel determinations [4, 5, 7, 42]. The application of AI to these databases would be the first step towards standardising objective tests to attain reliable sensorial quality assessments. In addition, AI techniques offer researchers the possibility to obtain practical consequences from their activities. For instance, meat researchers have developed the technology to assess the colour of meat using the CIE system [45]. The determination of the colour of meat by objective methods has been widely used [42, 46, 47;] but the relationship between these objective determinations and consumer acceptance of meat is not clear. This is an interesting problem because colour, along with tenderness, are the most important attributes when judging the quality of meat [43]. Probably, the further use of AI techniques in food science might be based on the possibility of food industry offering consumer products labelled according to their different sensorial qualities.

Acknowledgements

This paper was partially funded by a grant from CICYT-FEDER, No. 1FD97-1633 and by DGES PB98-1556. The authors would like to thank María Fernández, Jesús Fernández Tejón and Fernando Ureña from Empresa Asturiana de Servicios Agrarios (EASA) for their invaluable help and comments, and the authors of Cubist and M5 for allowing us to use their software.

References

1. Linko, S. (1998). Expert systems – what can they do for the food industry? *Trends in Food and Technology*, 9: 3-12.
2. Kempster, A. J. and Harrington, G. 1980. The value of “fat corrected” conformation as an indicator of beef carcass composition within and between breeds. *Livestock Prod. Sci.*, 7: 361-372.
3. Bourne, M.C., Food texture and viscosity: Concept and measurement, pp. 1-23 and 247-279, Academic Press, London 1982.
4. Casutt, M.M., Martin, Scheeder, R.L., Escher, F., Dufey, P.A. and Kreuzer, M., 1999. Relating texture properties and composition of bovine fat tissue. *Fett/Lipid*, 101 (8): 283-290.
5. Cierach, M. and Majewska, K., 1997. Comparison of instrumental and sensory evaluation of texture of cured and coked beef meat. *Nahrung* 41 (6): 366-369.
6. Guerrero, L. and Guàrdia, M.D., 1999. Evaluación de la ternera en carne de ternera: relación entre las medidas sensorial e instrumental según la dirección de las fibras musculares. *ITEA* (1999), Vol. Extra 20 (1), 11-14.

7. Mathoniere, C., Mioche, L., Dransfield, E., and Culioli, J. (2000). Meat texture characterisation: comparison of chewing patterns, sensory and mechanical measures. *Journal of Texture Studies* 31: 183-203
8. Goyache, F., del Coz J.J., Quevedo, J.R., López, S., Alonso, J., Ranilla, J., Luaces, O., Alvarez, I., Bahamonde, A. (2001): Using artificial intelligence to design and implement a morphological assessment system in beef cattle. *Animal Science* 73: 49-60.
9. Lincklaen Westenberg, H.W., de Jong, S., van Meel, D.A., Quadt, J.F.A., Backer, E. and Duin. R.P.W. (1989). Fuzzy set theory applied to product classification by a sensory panel", *Journal of Sensory Studies* 4: 55-72.
10. Nilsson, N.J. (1998). *Artificial Intelligence: a New Synthesis*. Morgan Kaufmann, San Francisco, CA, USA.
11. Rich, E., and Knigh, K. (1991). *Artificial Intelligence* (second edition). McGraw-Hill, New York, NY, USA.
12. Quinlan, J. R. (1993): *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.
13. Michalski, R.S., Bratko, I., and Kubat, M. (1998). *Machine Learning and Data Mining, Methods and Applications*. John Wiley and Sons Ltd., Chichester, England, UK.
14. Quinlan, J. R. (1993): Combining instance-based and model-based learning. En *Proceedings Tenth International Machine Learning Conference*. Amherst, MA: Morgan Kaufmann.
15. Quinlan, J. R. (1992): Learning with continuous classes. En *Proceedings 5th Australian Joint Conference on Artificial Intelligence*. World Scientific, Singapore, 343-348.
16. Wang Y. and Witten I.H. (1997): Inducing Model Trees for Continuous Classes. In *Poster Papers - 9th European Conf. on Machine Learning*, 128--137. Prague, Czech Republic
17. Cubist (2000): Release 1.09, <http://www.rulequest.com/cubist-info.html>
18. Quevedo, J. R. and Bahamonde, A. (1999): Aprendizaje de Funciones Usando Inducción sobre Clasificaciones Discretas. *Proceedings CAEPIA'99 VIII Conferencia de la Asociación Española para la Inteligencia Artificial*, Vol. I, 64-71. Murcia, Spain, 16-19 November 1999.
19. Aha, D.W., Kibler, D. and Albert, M.K. (1991): Instance-based learning algorithms. *Machine Learning*, 6, 37-66.

20. Widmer, G. (1993): Combining Knowledge-Based and Instance-Based Learning to Exploit Qualitative Knowledge. *Informatica* 17, Special Issue on Multistrategy Learning, 371-385.
21. Del Coz, J. J., Luaces, O., Quevedo, J.R., Alonso, J., Ranilla, J. and Bahamonde, A. (1999): Self-Organizing Cases to Find Paradigms. *Lecture Notes in Computer Sciences*, Springer-Verlag, Vol. 1606, 527-536.
22. Wettschereck, D., Aha, D.W. and Mohri, T. (1997) in 'A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms'. *Artificial Intelligence Review*, 11, pp. 273-314, 1997.
23. Blum, A. and Langley, P. (1997) in 'Selection of relevant features and examples in Machine Learning. In *Artificial Intelligence Journal*, special issue on relevance', Vol. 97, N 1-2, pp. 245-271.
24. John, G.H., Kohavi, R. and Pfleger, K. (1994) Irrelevant features and the subset selection problem. In *Proceedings Eleventh International Conference on Machine Learning*, pp. 121-129. New Brunswick, NJ (Morgan Kaufmann, San Mateo, CA). 1994.
25. Kira, K. and Rendell, L.A. 1992. A practical approach to feature selection. *Proceedings of the 9th International Conference on Machine Learning*, pp. 249-256. Aberdeen, Scotland: Morgan Kaufmann.
26. Schoonover, C.O., Brungardt, H.H., Carpenter, J.W., Guenther, J.J., King, G.T., Orts, F.A., Palmer, A.Z., Ramsey, C.B., Rust, R.E. and Zinn, D.W. 1968. Guides for beef carcass evaluation. *Recommended Guides for Carcass Evaluation and contest*. American Meat Science Association. 36 S, Wabash Avenue, Chicago, IL, USA.
27. Bass, J.J., Colomer-Rocher, F. and Johnson, D.L. 1981. Relationships between New Zealand beef conformations classes, carcass composition and muscle dimensions. *N. Z. J. Agric. Res.*, 24: 281-284.
28. De Boer H., Dumont B.L., Pomeroy R.W., Weniger J.H. 1974. Manual on E.A.A.P. reference methods for the assessment of carcass characteristics in cattle. *Livestock Prod. Sci.*, 1, 151-164.
29. Colomer-Rocher, F. 1992. El carácter conformación en los bovinos y sus determinantes biológicos. *Proceedings of Jornadas sobre tecnología de valoración de canales y carnes y defensa de la calidad de los productos ganaderos*. Feria Internacional Ganadera Quinto Centenario. Zafra, Badajoz, España, Mayo.
30. Cross, H.R., Gilliland, D.A., Durland, P.R. and Seildman, S. (1983). Beef carcass evaluation by use of a video image analysis system. *J Anim Sci*, 57: 908-917.
31. Wood, J.D., Newman, P.B., Miles, C.A. and Fisher, A.V. (1991). Video image analysis: comparisons with other novel techniques for carcass assessments. *Proceedings of a symposium on the electronic evaluation of meat in support of value-based marketing*, March 27-28, Purdue University, In pp. 145-169.
32. Yeates, N.T.M. 1967. *Avances en Zootecnia*. Editorial ACRIBIA, Zaragoza, España, pp. 202.
33. Díez, J.; Alonso, J.; López, S.; Bahamonde, A.; Goyache, F. (2001). Una aplicación informática para la representación informática de la conformación de canales bovinas. *ITEA*, 22 Vol. Extra II: 550: 552.

34. Kempster, A.J. 1986. Estimation of the carcass composition of different cattle breeds and crosses from conformation assessments adjusted for fatness. *J Agric sci Camb*, 106: 239-254.
35. Luaces, O., del Coz, J. J., Quevedo, J.R., Alonso, J., Ranilla, J. Bahamonde, A. 1999. Autonomous Clustering for Machine Learning. *Lecture Notes in Computer Sciences*, LNCS No. 1606, pp. 497-506, Springer-Verlag, Berlín.
36. Bahamonde, A., de la Cal, E. A., Ranilla, J., Alonso, J. (1997): Self-organizing symbolic learned rules. *Lecture Notes in Computer Science*, LNCS N° 1240, 536-545. Springer-Verlag. Berlin.
37. González Perez, F.; Alonso González, J. (2001) Aprendizaje de categorías continuas a partir de ejemplos. *Proceedings CAEPIA 2001 IX Conferencia de la Asociación Española para la Inteligencia Artificial*, Vol. I, 3-12. Gijón, Spain, November 2001
38. Domingos, P. (1996). Unifying instance-based and rule-based induction. *Machine Learning*, 24, 141-168.
39. Ranilla, J., Mones, R., & Bahamonde, A. (1998). El Nivel de Impureza de una regla de clasificación aprendida a partir de ejemplos. *Revista Iberoamericana de Inteligencia Artificial*, 4, 4-11.
40. Honikel, K.O. (1998). Reference methods for the assessment of physical characteristics of meat. *Meat Science*, 49: 447-457.
41. Wood, J.D., Holder, J.S. And Main D.C.J. (1998) Quality assurances schemes. *Proceedings of 44th ICoMST*, Barcelona, Spain. pp. 206-215.
42. Sañudo, C., Olleta, J.L., Campo, M.M., Panea, B., Renand, B., Turin, F., Jabet, S., Osoro, K., Oliván, C., Noval, G., García, M.J., García, D., Cruz-Sagredo, R., Oliver, M.A., Gil, M., Gispert, M., Serra, X., Guerrero, L., Espejo, M., García, S., López, M., Izquierdo, M., Quintanilla, R., Martín, M. and Piedrafita, J. (2000). Meat quality of ten cattle breeds of the Southwest of Europe. Project FAIR CT95 0702. Final report, pp. 190-231.
43. Love, J. (1994). Product acceptability evaluation. In *Quality attributes and their measurement in meat, poultry and fish products*. Eds: A.M. Pearson and T.R. Dutson. Pp. 337-358. Blackie Academic and Professional. Glasgow, Scotland, UK.
44. Issanchou, S. (1996). Consumer expectations and perceptions of meat and meat product quality. *Meat Science*, 43 (S): 5-19.
45. Judd, D. B. (1952). *Colour in Business, Science and Industry*. John Willey and sons, New York, NY, USA.
46. Strange, E.D., Benedict, R.C., Gugger, R.E., Metzger, V.G. and C.E. Swift (1974). Simplified methodology for measuring meat color. *J Food Science*, 39: 988-992.

47. Eagerman, B.A., Clydesdale, F.M. and Francis F.J. (1977). Determination of fresh meat color by objective methods. *J Food Science*, 42: 707-724.

Figure 1.- Graph (a) represents a variable showing a clearly linear behaviour, with $r=0.81$. The variables represented in Graph (b) 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.

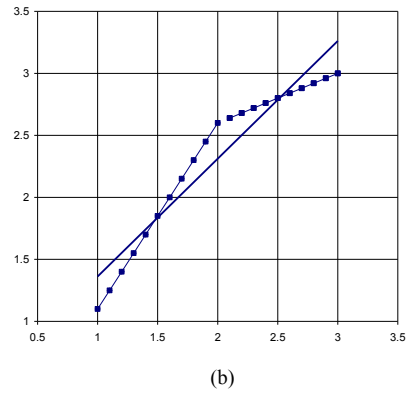
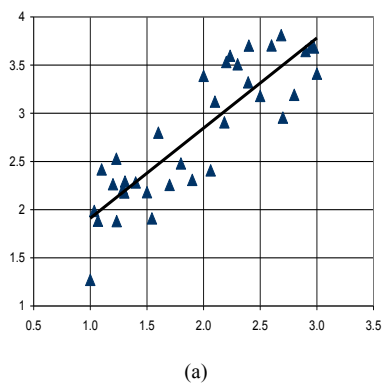


Table 1: Description of the 28 attributes used to represent bovine carcass conformation grading in 93 carcasses building the training set and correlation coefficients with a conformation score for each attribute. Attributes are described in cm for distances between two given key points, cm² for areas and cm³ for volumes; profiles and ratios are described in the corresponding units; carcass weight is in Kg.

Attribute	Mean	Std Dev	Minimum	Maximum	Correlation with conformation score
Carcass weight	260.73	72.50	128.00	560.00	0.24*
Thigh width	29.19	6.89	22.09	87.65	0.13
Thigh length	45.88	3.41	37.36	56.02	-0.04
Thigh ratio (width/length)	0.63	0.16	0.37	1.99	0.11
Hindlimb length	66.13	6.28	35.64	83.69	-0.31**
Shoulder length	41.93	4.98	19.86	56.05	-0.21*
Shoulder width	36.71	10.46	23.76	126.30	-0.11
Lateral shoulder height	4.87	2.00	1.25	17.67	0.28**
Shoulder area	1531.62	515.32	140.37	5300.30	-0.2
Shoulder volume	8210.41	9470.57	1708.82	93697.00	0.02
Round profile	-1.98E-02	8.61E-03	-5.30E-02	1.47E-02	-0.22*
Shoulder profile	-1.66E-02	9.93E-03	-3.68E-02	2.56E-02	-0.15
Topside profile	-2.36E-02	1.04E-02	-3.80E-02	3.55E-02	-0.32**
Back length	58.30	16.26	31.74	173.90	-0.16
Front width of back	11.97	4.09	4.65	30.74	-0.11
Rear width of back	12.97	4.36	4.84	39.19	0
Back width ratio (front/rear)	0.91	0.29	-0.81	1.69	-0.1
Lateral back width at hip	13.94	2.68	8.46	21.24	0.09
Lateral back width at chest	15.01	2.74	9.31	22.87	-0.04
Lateral back width ratio(hip/chest)	0.93	0.19	0.09	1.39	0.24*
Back volume	11655.00	15879.00	1443.98	139905.00	-0.02
Hips profile	-3.05E-02	4.95E-02	-1.30E-01	1.66E-01	0.14
Thigh profile	-2.61E-02	2.03E-02	-2.04E-01	-7.00E-04	-0.32**
Neck profile	8.36E-02	9.30E-01	-5.63E-02	8.95E+00	0.11
Carcass length	114.47	13.78	63.15	157.10	-0.26*
Chest depth	43.98	5.83	23.34	61.03	-0.36***
Belly depth	33.41	5.54	16.42	52.86	-0.08
Blockiness index	2.28	0.59	1.31	4.11	0.35***
Conformation score	3.80	1.14	0.75	5.25	

*, ** and *** mean significant correlation coefficients for $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively

SERIDA 11/27/01 10:51 A.M.

Eliminado:

Table 2.- Formulas obtained by simple linear regression and by SAFE [18] to assess SEUROP classification for bovine carcasses. The training examples set has 243 elements and is analysed in the rest of the paper.

Simple linear regression

Used by 243 examples. Average error when used = 0.604983,

$$\text{Score} = -0.00492146 \cdot \text{att26} - 0.40605 \cdot \text{att28} - 38.1857 \cdot \text{att4} - 34.2187 \cdot \text{att3} + 2.95894$$

SAFE's output

Regression rule # 1 used by 24 examples. Average error when used = 0.0138889,

$$\text{Score} = +1372.87 \cdot \text{att4} + 5.03689 \cdot \text{att26} + 823.185 \cdot \text{att3} - 62.427 \cdot \text{att28} - 89.9846$$

$$\text{If } \text{att26} \in [29.656, 33.7001] \ \& \ \text{att28} \in [0.323309, 0.505332] \ \& \ \text{att3} \in [-0.020545, -0.007526] \ \& \ \text{att4} \in [-0.027978, -0.013209]$$

Regression rule # 2 used by 11 examples. Average error when used = 0.11675,

$$\text{Score} = +57.6305 \cdot \text{att28} - 55.7315 \cdot \text{att4} - 343.914 \cdot \text{att3} - 27.9049$$

$$\text{If } \text{att28} \in [0.350467, 0.389955] \ \& \ \text{att26} \in [30.3177, 33.9848] \ \& \ \text{att3} \in [-0.031378, -0.019873]$$

Regression rule # 3 used by 13 examples. Average error when used = 0.173077,

$$\text{Score} = +0.0820413 \cdot \text{att26} + 698.677 \cdot \text{att3} - 1306.89 \cdot \text{att4} - 15.291$$

$$\text{If } \text{att26} \in [27.5343, 28.9624] \ \& \ \text{att28} \in [0.503918, 0.525912] \ \& \ \text{att3} \in [-0.034231, -0.021452] \ \& \ \text{att4} \in [-0.032044, -0.02422]$$

Regression rule # 4 used by 10 examples. Average error when used = 0.160846,

$$\text{Score} = -14.4474 \cdot \text{att4} + 4.38894$$

$$\text{If } \text{att28} \in [0.342866, 0.412822] \ \& \ \text{att26} \in [29.6045, 39.1749] \ \& \ \text{att3} \in [-0.037217, -0.018723] \ \& \ \text{att4} \in [-0.036658, -0.023106]$$

Regression rule # 5 used by 9 (9 Orig.) examples. Average error when used = 0.25399,

$$\text{Score} = -31.2425 \cdot \text{att4} + 3.43667$$

$$\text{If } \text{att3} \in [-0.031478, -0.023376] \ \& \ \text{att26} \in [16.4233, 22.7168]$$

Regression rule # 6 used by 49 examples. Average error when used = 0.460772,

$$\text{Score} = +4.34373 \cdot \text{att28} + 3.45121$$

$$\text{If } \text{att4} \in [-0.038823, -0.012467] \ \& \ \text{att26} \in [16.4233, 40.9314] \ \& \ \text{att28} \in [0.0, 0.367619] \ \& \ \text{att3} \in [-0.037217, -0.005866]$$

Regression rule # 7 used by 47 examples. Average error when used = 0.532492,

$$\text{Score} = -92.1068 \cdot \text{att3} + 2.18215$$

$$\text{If } \text{att4} \in [-0.035248, -0.01419] \ \& \ \text{att28} \in [0.420473, 0.53389] \ \& \ \text{att26} \in [27.1925, 35.0929] \ \& \ \text{att3} \in [-0.034231, -0.014413]$$

Regression rule # 8 used by 52 examples. Average error when used = 0.57655,

$$\text{Score} = -1.30004 \cdot \text{att28} - 13.9236 \cdot \text{att4} - 27.2842 \cdot \text{att3} + 4.04745$$

$$\text{If } \text{att26} \in [16.4233, 45.2727] \ \& \ \text{att3} \in [-0.037979, 0.014183] \ \& \ \text{att4} \in [-0.038167, -0.015889]$$

Regression rule # 9 used by 15 examples. Average error when used = 0.876216,

$$\text{Score} = -34.3474 \cdot \text{att3} - 37.9177 \cdot \text{att4} + 2.62835$$

$$\text{If } \text{att4} \in [-0.036677, -0.000742] \ \& \ \text{att3} \in [-0.038059, 0.014183]$$

Figure 2.- Each beef carcass was photographed in three positions: lateral (a), medial (b) and dorsal (c). An operator then manually marked a set of relevant points and curves; only 8 points are really needed, as will be shown later. To represent the profile convexities, we isolate the curve arcs (d) that borders them as a real function (e).

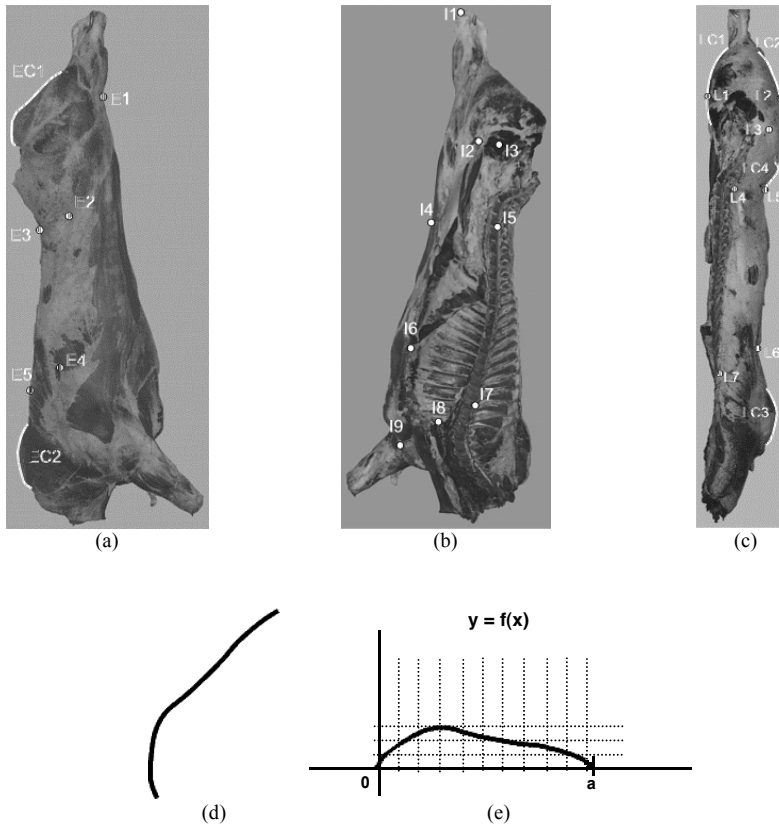


Table 3.- Average absolute errors of learning systems obtained by cross validation of 10 folders repeated 5 times over the training set described in the text. Notice that the average absolute difference of our expert classifiers with the average SEURO scores of the 3 is: 0.41 for the first expert, 0.39 for the second and 0.41 for the third. The system mean that always (i.e. unconditionally) returns the average of the scores of the training set has a cross validation error of 0.74. Additionally, the number of functions or prototypes selected to make predictions on unseen cases accompanies the average absolute errors for each learning system.

Number of attributes	SAFE		BETS		LC3		Cubist		M5		Linear Regression
	Err.	#fun	Err.	#prot	Err.	#prot.	Err.	#fun	Err.	#fun	Err.
28	0.43	5.0	0.46	37.4	0.43	53.2	0.44	8.74	0.43	15.96	0.45
15	0.44	7.64	0.43	34.48	0.42	34.92	0.43	7.04	0.40	15.96	0.44
4	0.51	7.94	0.45	33.82	0.44	34.76	0.45	7.82	0.49	16.16	0.62