ELSEVIER

Identifying market segments in beef: Breed, slaughter weight and ageing time implications

J. Díez [a,*], J.J. del Coz [a], A. Bahamonde [a], C. Sañudo [b], J.L. Olleta [b], S. Macie [b],
M.M. Campo [b], B. Panea [c], P. Albertí [c]

[a] *Artificial Intelligence Centre, University of Oviedo, 33271 Gijón, Spain*
[b] *Veterinary Faculty, University of Zaragoza, 50013 Zaragoza, Spain*
[c] *Centre of Agricultural and Technological Research, 50080 Zaragoza, Spain*

Abstract

In this paper we propose a method to learn the reasons why groups of consumers prefer some beef products to others. We emphasise the role of groups since, from a practical point of view, they may represent market segments that demand different products. Our method starts representing people's preferences in a metric space; there we are able to define a kernel based similarity function that allows a clustering algorithm to identify significant groups of consumers with homogeneous likes. Finally, in each cluster, we developed, with a support vector machine (SVM), a function that explains the preferences of those consumers grouped in the cluster. The method was applied to a real case of consumers of beef that tasted beef from seven Spanish breeds, slaughtered at two different weights and aged for three different ageing periods. Two different clusters of consumers were identified for acceptability and tenderness, but not for flavour. Those clusters ranked two very different breeds (Asturiana and Retinta) in opposite order. In acceptability, ageing period was appreciated in a different way. However, in tenderness most consumers preferred long ageing periods and heavier to lighter animals.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Consumers; Preferences; Clustering; Acceptability; Machine learning; Artificial intelligence

1. Introduction

Studies about consumers' preferences are appreciated by the food industry since they can explain consumers' decisions (Verbeke & Vackier, 2004). It is the hedonic value of the product that could determinate the future purchasing decisions, with the industry seeking authenticity or quality reassurance. This is especially true if the information or signals given on the labels are shown and easily identifiable in the market, at least, for people who are concerned about meat.

The preferences for food products address the strategies of industries and breeders, and should be carefully considered when export and commercial policies are designed. In this paper we present a method to deal with data collected from panels of consumers in order to discover groups with differentiated tastes; these groups may constitute significant market segments that demand different kinds of food products. Additionally, our approach studies the factors that could contribute to the success or failure of different meat types in each segment.

From a conceptual point of view, consumer panels are made up of untrained people; these are asked to rate their degree of acceptance or satisfaction about the tested products on a scale. The aim is to be able to relate product descriptions (human and mechanical) with consumer preferences. Simple statistical methods cannot cope with this task. In fact, this is not a straightforward task; the reason is that when we are aiming to induce a function that maps object descriptions into ratings, we must consider that the consumers' ratings are just a way to express their relative

---

* Corresponding author.
  E-mail address: jdiez@aic.uniovi.es (J. Díez).

preferences about the products presented in the same testing session. Additionally, it is necessary to realise that numerical ratings do not mean the same for all the people, the scales used may be quite different. Discussions about ratings and preferences can be found in the study of Cohen, Shapire, and Singer (1999), in general, and in those of Díez et al. (2003), Del Coz et al. (2004) or Luaces et al. (2004) in the context of food.

From a practical point of view, the market is not interested in preferences of individual consumers, the purpose of marketing studies of sensorial data is to discover, if they exist, widespread ways to appreciate food products that can be considered as market segments. These segments can be seen as *clusters* of consumers with similar tastes. In this paper, we will show that the similarity of preference criteria of consumers can be computed in a space of preference functions by means of a kernel-based method.

Beef quality is influenced by many intrinsic and extrinsic factors, which have been widely studied. However, many efforts are still required to include the different biological or productive possibilities and technological variations together with their respective interactions and, at the end, their influence on meat preferences. Thus, the effect of the breed per se on meat characteristics is not so clear; differences on meat quality could depend on sex, slaughter weight and/or ageing (Monsón, Sañudo, & Sierra, 2005). On the other hand, ageing is widely accepted as a main factor on meat quality and consumer perception. In this way, the clustering method presented in this paper was used to process a data set that collects the ratings of a panel of beef consumers where the 171 panellists tasted beef meat (Sañudo et al., 2004) from different breeds, live weights and ageing periods.

2. Material and methods

2.1. The general approach

The main assumption behind the approach presented in this paper is that we are able to map people's preferences into a metric space in such a way that we can assume some kind of continuity. A first attempt to provide such a mapping would consist in associating, to each consumer, the vector of his or her ratings, taking the set of samples as indexes. However, this is not a wise option since ratings have only a relative meaning, and therefore they cannot assume an absolute role. There is a *batch effect*: a product will obtain a higher/lower rating when assessed together with other products that are clearly worse/better. In fact, if we try to deal with sensory data as a regression problem, we will fail (Díez et al., 2004); due to this batch effect, the ratings have no numerical meaning: they are only a relative way to express preferences between products of the same session.

To overcome this, instead of ratings, we can assign to each product its ordinal position in the ranking of preferences. Unfortunately, this is not always possible since, in general, the size of the sample of food prevents consumers from tasting all products. Hence, we cannot ask people to spend long periods rating the whole set of food samples. Typically, each consumer only participates in one or a small number of tasting sessions, usually in the same day. Notice that tasting a large sample of food may be physically impossible, or the number of tests performed would damage the sensory capacity of consumers.

The consequence is that consumers' rankings are not comparable because they deal with different sets of products. Thus, in this case we will codify people preferences by the weighting vector of a linear function (called *preference* or *ranking function*) in a high dimensional space: the space of features where we finally represent the descriptions of food products. Then, the similarity is defined by means of the kernel attached to the representation map in a sense that we will explain in detail in the following. The computational aspects of this approach can be seen in Díez, Del Coz, Sañudo, Albertí, and Bahamonde (2005).

Once we have people preferences represented in a metric space, and we have defined a similarity function, then we use a clustering algorithm. Although there are other possibilities, we used the nonparametric hierarchical clustering algorithm (Dubnov et al., 2002) that uses a proximity matrix of pairwise relations that directly captures the intention of the similarity functions. Then, we only need to explain the meaning and implications of each cluster in the context of food products. For this purpose, we will learn a preference or ranking function from the union of preference judgements expressed by members of the cluster; this will provide the consensus assessment function of the cluster.

2.2. Description of the beef experiment

The clustering method presented in this paper was used to process a database described in Sañudo et al. (2004). Also, more information about these breed characteristics and procedures could be found in other papers (Campo, Sañudo, Panea, Albertí, & Satolaria, 1999; Campo et al., 2000; Gil et al., 2001; Insausti, Berian, Alzueta, Carr, & Purroy, 2004).

For this experiment, 103 male calves from seven Spanish breeds were housed, after having been collected from different farms at an initial live weight between 240 and 290 kg. Animals were fed intensively with concentrate and cereal straw ad libitum and slaughtered at a EU licensed abattoir, following the normal commercial practice, to obtain two type of carcasses: light carcasses, from animals with a live weight between 300 and 350 kg (light); and heavy carcasses, from animals at 530–560 kg (heavy). Breeds and weights were uniformly distributed in the set of animals. Additionally, to test the influence of ageing in consumers' appreciation, each steak (*Longissimus dorsi* muscle) was aged for three different periods: 1, 7, and 21 days.

On the other hand, the seven breeds that we have used constitute a wide representation of beef cattle. These breeds

can be grouped into four types (Campo et al., 1999, 2000): double muscled (DM, one breed), fast growth (FG, two breeds), dual purpose (DP, one breed), and unimproved type (UT, three breeds). Tables 1 and 2 show the carcass composition characteristics for light and heavy animals of each breed, respectively. In the experimental results reported in Section 3 we used the carcass composition of each animal; the breed was only used to provide an explanation of the preferences of clusters.

The muscle *longissimus dorsi* was cut into 2-cm steaks at 24 h post-mortem. Steaks were vacuum packaged and randomly allocated to each ageing time. Samples that were aged for 1 day were immediately frozen and the rest were kept at 2–4 °C for 7 and 21 days, then frozen and stored at −18 °C until analysis. Steaks were thawed in their vacuum bags in tap water for 4 h before each session to an internal temperature of 17–19 °C. Samples were cooked in a double plate grill, preheated at 200 °C, until they reached an internal temperature of 70 °C, which was monitored by an internal thermocouple (Jenway 2000). Meat was trimmed of any external connective tissue, cut into approximately $2 \times 2 \times 2$-cm samples, wrapped in coded aluminium foil and stored in warm pans until testing. Consumer tests were performed in a controlled sensory analysis laboratory (ISO 8589, 1988), with individual booths under red lighting to mask differences in meat colour.

The panel was organised in a set of tasting sessions, where a group of consumers assessed the available beef. Each consumer (393 consumers in total) participated in only one session (40 in total), that included two sub-sessions, evaluating in total 10 samples: three ageing times of the same breed and

the seven breeds at the same ageing time. The same set of samples was presented in a different order to each consumer inside the same sub-session, following a randomised design for order and carry over effects. Sessions were performed by groups of a maximum of 10 consumers and a minimum of 9. Before starting each session consumers were asked to fill in some personal data (Table 3). The consumers were all surveyed during the entire assay to make sure it was carried out

Table 3
Socio-demographic characteristics of the consumers ($n = 393$)

Characteristics	Percentage (%)
Sex	
Male	46.6
Female	51.7
Missing information	1.7
Age	
<25 years	24.9
26–35 years	20.7
36–45 years	21.2
46–55 years	15.6
>56 years	16.2
Missing information	1.4
Profession	
Student	22.2
Worker	7.9
Technique	18.1
High graduate	20.4
Teacher – research	10.3
Pensioner	5.3
Housewife	13.6
Missing information	2.2

Table 1
Average percentages 10th rib dissection and intramuscular fat content for light carcasses of each breed used in the experiment

Breed		Fat (%)		Bone (%)	Muscle (%)	Intramuscular fat (%)
Name	Type	Intermuscular	Subcutaneous			
Asturiana Valles	DM	4.78[b]	0.96[de]	16.35[c]	77.91[a]	0.95[b]
Avileña	UR	11.14[a]	2.55[bc]	19.92[b]	66.39[c]	1.60[ab]
Morucha	UR	11.18[a]	3.15[ab]	19.44[b]	66.23[c]	2.03[a]
Parda Alpina	DP	7.54[b]	1.51[cde]	21.73[ab]	69.22[bc]	1.13[b]
Pirenaica	FG	7.50[b]	2.16[bcd]	19.79[b]	70.54[b]	1.20[b]
Retinta	UR	11.71[a]	3.90[a]	23.35[a]	61.05[d]	1.59[ab]
Rubia Gallega	FG	4.93[b]	0.87[e]	16.93[c]	77.27[a]	0.93[b]

Table 2
Average percentages 10th rib dissection and intramuscular fat content for heavy carcasses of each breed used in the experiment

Breed		Fat (%)		Bone (%)	Muscle (%)	Intramuscular fat (%)
Name	Type	Intermuscular	Subcutaneous			
Asturiana Valles	DM	4.77[d]	0.79[d]	15.53[cd]	78.91[a]	0.83[c]
Avileña	UR	14.94[ab]	4.38[ab]	18.67[ab]	62.01[cd]	2.88[a]
Morucha	UR	13.93[abc]	3.81[ab]	19.10[ab]	63.16[cd]	2.19[ab]
Parda Alpina	DP	11.76[bc]	3.12[bc]	19.99[a]	65.12[c]	2.51[a]
Pirenaica	FG	10.76[c]	3.99[ab]	14.52[d]	70.73[b]	1.80[ab]
Retinta	UR	16.61[a]	5.60[a]	18.44[abc]	59.34[d]	2.67[a]
Rubia Gallega	FG	6.67[d]	1.57[cd]	16.12[bcd]	75.64[a]	1.34[bc]

properly. In the clustering method presented here we only used the rates of 171 consumers (out of 393) since the rest did not exhibit enough discrimination power in a sense that will be clear below.

Consumers evaluated three different attributes for each sample: tenderness, flavour quality and overall acceptability using a structured category scale of 10 points. In the scale, 1 was labelled as extremely tough, extremely bad flavour and extremely bad overall acceptability, and 10 was labelled as extremely tender, extremely good flavour quality and extremely good overall acceptability.

2.3. Vectorial representation of preference criteria

As explained above, in order to compare the preference criteria of consumers we need to state a common language. We cannot use the ratings assigned by the consumers, since they have rated different sets of samples. Thus, we are going to induce a reasonable extension of the preferences expressed by each consumer to obtain a function able to capture the pairwise orderings, not the rates. Then we will manage to define similarities in the space of those functions.

Although there are other approaches to learn preferences, we will follow those described in Herbrich, Graepel, and Obermayer (2000), Joachims (2002) and Bahamonde et al. (2004). Then, we will try to induce a real *preference*, *ranking* or *utility function* '*f*' from the input space of object descriptions, say R^d, in such a way that it maximises the probability of having $f(v) > f(u)$ whenever v is preferable to u; we call such pairs, *preference judgments*.

After having a set of ratings given by a consumer '*c*', we take into account the session where the ratings have been assessed (Del Coz et al., 2004; Luaces et al., 2004), as was explained in Section 2.1. Thus, for each session we include in the set of preference judgments, PJ_c, the pairs (v, u) whenever consumer c assessed to the sample represented by v a higher rating than to the sample represented by u.

In order to induce the preference or ranking function (Herbrich et al., 2000), we are going to use a support vector machine (SVM). These *machines*, when dealing with binary classification tasks, are algorithms that return a function whose graph maximises the separation margin between the classes; the classification of cases is given by the sign returned by the function. To induce such functions, SVMs use well known and founded techniques of Quadratic Programming; see for instance the book of Shawe-Taylor and Cristianini (2004). A very interesting property of SVMs is that they do no use the coordinates of input vectors, instead they only use scalar products of pairs of inputs. The practical consequence of this fact is the so-called kernel-trick that allow us to use inputs $\mathbf{x} \in R^d$ represented by $\phi(\mathbf{x})$ in a high dimensional space (called the feature space), as for instance, the space of monomials of d variables. The reason is that it is easy to prove that the scalar product of a couple of transformed vectors is given by

$$K(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle = (\langle \mathbf{x}, \mathbf{y} \rangle + 1)^g \tag{1}$$

Functions like K of Eq. (1) are called kernel functions. The function of Eq. (1) is called the polynomial kernel of degree g. If $K(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle$ we have the linear kernel.

In the case of beef meat, let us assume that each piece of meat tasted by consumers is represented by $\phi(\mathbf{x})$ where $\mathbf{x}$ is the vectorial arrangement of data that identifies the sample. We assume also that we can compute the scalar products by Eq. (1). In this framework, given the set of preference judgments of a consumer c, PJ_c, we define a set of binary classification training set as follows:

$$E_c = \{(v, u, +1), (u, v, -1) : (v, u) \in PJ_c\} \tag{2}$$

So, notice that now inputs are pairs of pieces of meat. Then we define a new transformation by

$$\Psi(v, u) = \phi(v) - \phi(u) \tag{3}$$

Therefore, the associated kernel to this transformation (called the Herbrich's kernel attached to K of Eq. (1)) is given by

$$
\begin{aligned}
&\mathbf{K}\big(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}; \mathbf{x}^{(3)}, \mathbf{x}^{(4)}\big) \\
&= \big\langle \Psi\big(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}\big), \Psi\big(\mathbf{x}^{(3)}, \mathbf{x}^{(4)}\big) \big\rangle \\
&= \big\langle \big(\phi\big(\mathbf{x}^{(1)}\big) - \phi\big(\mathbf{x}^{(2)}\big)\big), \big(\phi\big(\mathbf{x}^{(3)}\big) - \phi\big(\mathbf{x}^{(4)}\big)\big) \big\rangle \\
&= \big\langle \phi\big(\mathbf{x}^{(1)}\big), \phi\big(\mathbf{x}^{(3)}\big) \big\rangle - \big\langle \phi\big(\mathbf{x}^{(1)}\big), \phi\big(\mathbf{x}^{(4)}\big) \big\rangle \\
&\quad - \big\langle \phi\big(\mathbf{x}^{(2)}\big), \phi\big(\mathbf{x}^{(3)}\big) \big\rangle + \big\langle \phi\big(\mathbf{x}^{(2)}\big), \phi\big(\mathbf{x}^{(4)}\big) \big\rangle \\
&= K\big(\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\big) - K\big(\mathbf{x}^{(1)}, \mathbf{x}^{(4)}\big) \\
&\quad - K\big(\mathbf{x}^{(2)}, \mathbf{x}^{(3)}\big) + K\big(\mathbf{x}^{(2)}, \mathbf{x}^{(4)}\big)
\end{aligned}
\tag{4}
$$

Finally, the separation function induced by a classification SVM from E_c (Eq. (2)) with kernel $\mathbf{K}$ will be a function $\mathbf{F}$: $R^d \times R^d \to R$ of the form

$$
\begin{aligned}
\mathbf{F}(\mathbf{x}, \mathbf{y}) &= \sum_{s \in S(c)} \alpha_s z_s \big\langle \phi\big(\mathbf{x}_s^{(1)}\big) - \phi\big(\mathbf{x}_s^{(2)}\big), \phi(\mathbf{x}) - \phi(\mathbf{y}) \big\rangle \\
&= \sum_{s \in S(c)} \alpha_s z_s \mathbf{K}\big(\mathbf{x}_s^{(1)}, \mathbf{x}_s^{(2)}, \mathbf{x}, \mathbf{y}\big)
\end{aligned}
\tag{5}
$$

where α_s are some weighting coefficients, z_s stands for the class +1 or −1, and $S(c)$ is the set of support vectors: those whose corresponding α_s are not zero. In fact, it is not necessary to consider the examples of class '−1' (see Eq. (2)) if we are able to ensure that the SVM is going to search for a separating hyperplane between those that pass through the origin of coordinates. This can be done, for instance, if we use the implementation SVM^{light} of Joachims (1999, chap. 11).

Then it can be shown that a ranking or preference or utility function f_c is given by

$$
\begin{aligned}
f_c(x) &= \sum_{s \in S(c)} \alpha_s z_s \big\langle \phi\big(\mathbf{x}_s^{(1)}\big) - \phi\big(\mathbf{x}_s^{(2)}\big), \phi(\mathbf{x}) \big\rangle \\
&= \sum_{s \in S(c)} \alpha_s z_s \big(k\big(\mathbf{x}_s^{(1)}, \mathbf{x}\big) - k\big(\mathbf{x}_s^{(2)}, \mathbf{x}\big)\big)
\end{aligned}
\tag{6}
$$

Therefore, f_c can be represented by some weight vector w^c in the high dimensional space of features such that

$$f_c(x) = \langle w^c, \phi(x) \rangle. \tag{7}$$

In symbols

$$\mathbf{w}^c = \sum_{s \in S(c)} \alpha_s z_s \big(\phi(\mathbf{x}_s^{(1)}) - \phi(\mathbf{x}_s^{(2)}) \big) \tag{8}$$

Notice that (6) defines the ranking of an object represented by a vector x. This is not an absolute value; its importance is the relative position that gives to x against to the other objects y in the *competition* for gaining the appreciation of consumer c.

Now we only need to define the distance of consumers' preferences. Given that preferences are codified by those weighting vectors, we define the similarity of the preferences of consumer c and c' by the cosine of their weighting vectors. In symbols

$$\text{similarity}(\mathbf{w}^c, \mathbf{w}^{c'}) = \cos(\mathbf{w}^c, \mathbf{w}^{c'}) = \frac{\langle \mathbf{w}^c, \mathbf{w}^{c'} \rangle}{\|\mathbf{w}^c\| * \|\mathbf{w}^{c'}\|} \tag{9}$$

Given that this definition uses scalar products instead of coordinates of weighting vectors, we can easily rewrite (9) in terms of the kernels used in the previous derivations. The essential equality is:

$$\langle \mathbf{w}^c, \mathbf{w}^{c'} \rangle = \sum_{s \in S(c)} \sum_{l \in S(c')} \alpha_s \alpha_l z_s z_l \left\langle \phi(\mathbf{x}_s^{(1)}) - \phi(\mathbf{x}_s^{(2)}), \phi(\mathbf{x}_l^{(1)}) - \phi(\mathbf{x}_l^{(2)}) \right\rangle$$
$$= \sum_{s \in S(c)} \sum_{l \in S(c')} \alpha_s \alpha_l z_s z_l \mathbf{K}(\mathbf{x}_s^{(1)}, \mathbf{x}_s^{(2)}, \mathbf{x}_l^{(1)}, \mathbf{x}_l^{(2)}) \tag{10}$$

2.4. Clustering consumers with homogeneous tastes

In the previous section, we have associated one data point for each consumer in the space of preference criteria represented by ranking or preference functions. Moreover, we have defined a reasonable similarity measure for preference criteria; now we proceed to look for clusters of consumers with homogeneous tastes. For this purpose, we applied a nonparametric pairwise algorithm (Dubnov et al., 2002).

2.4.1. The clustering algorithm

Let $S = (s_{ij})$ be a square matrix where s_{ij} stands for the similarity between data points i and j; in our case, data points are the vectorial representation of the preference criteria of consumers, and similarities are given by Eq. (9). From now onwards, S will be called the *proximity matrix*. Then, matrix S is transformed iteratively, following a two step procedure that converges into a two value matrix (1 and 0), yielding a bipartition of the data set into two clusters. Then, recursively, the partition mechanism is applied to each of the resulting clusters represented by their corresponding submatrices. To guarantee that only meaningful splits take places, Dubnov et al. (2002) provide a cross-validation method that measures an index that can be read as a significance level; we will only accept splits whose level is above 0.90.

The basic iterative transformation uses the following formulae to go from iteration t to $t+1$:

$$p_{ij}(t+1) = \frac{s_{ij}(t)}{\max\{|s_{ik}(t)| : k\}}$$
$$s_{ij}(t+1) = \frac{1}{2} \left(\begin{array}{l} \sum_k p_{ik}(t+1) \log \frac{p_{ik}(t+1)}{\frac{1}{2}(p_{ik}(t+1)+p_{jk}(t+1))} \\ + \sum_k p_{jk}(t+1) \log \frac{p_{jk}(t+1)}{\frac{1}{2}(p_{jk}(t+1)+p_{ik}(t+1))} \end{array} \right) \tag{11}$$

The first step normalises the columns of the proximity matrix using the L_∞ norm; then the proximities are re-estimated using the Jensen–Shannon divergence. The idea is to formalise that the two preference criteria are close (after these two steps) if they were both similar and dissimilar to analogous sets of criteria before the transformation.

Notice that we could have used other available bottom-up clustering method; however, these methods tend to produce many clusters, and in sensory analysis applications, we do not expect that many market segments exists. Therefore, a top-down clustering (Dubnov et al., 2002) is more appropriate, but probably any other algorithm of this kind would yield similar results, although with the cost of using more customizing parameters.

2.4.2. What do consumers' clusters mean?

Given a set of clusters {Cluster(j): $j = 1$:n}, we have to explain the reasons that make people in each cluster have those similar criteria that make them different from people of other clusters. The best way to achieve this is to induce a preference function using product descriptions. The learning algorithm is the SVM explained in Section 3, but notice that now instead of using the preference judgments PJ_c sets of individual consumers, we consider for each cluster the union

$$PJ_{\text{cluster}(j)} = \bigcup_{c \in \text{cluster}(j)} PJ_c \tag{12}$$

The preference functions (see Eq. (7)) will be useful for two different things. Firstly, we can compute the average ranking of the cluster, and the estimation of the ranking position of future products given their descriptions. Secondly, it would be possible to determine the influence of each feature that describes food products in the acceptability by consumers of the market segment represented by clusters. Therefore, we would be able to design policies to improve the acceptability by different kinds of consumers.

3. Results and discussion

In this section, we report the outputs obtained with the database of beef consumers described in Section 2.2. Since the aim was to investigate the possibilities of existence of different sensibilities that could be understood as market segments, we were interested in widespread consumers' behaviour. Therefore, we first selected those people involved in our consumers' panel whose ratings gave rise to more than 30 preference judgments; these yielded a set

of 171 panellists. Notice that this is just a subset of all available panellists since samples rated with the same rate did not produce preference judgement pairs. In any case, this individual treatment contrasts with the global one that we previously reported (Del Coz et al., 2004; Díez et al., 2004; Luaces et al., 2004). In all these cases, we were interested in modelling the general opinion of consumers; so for each session, to summarise the opinions of consumers, we computed the mean of the ratings obtained by each piece of meat.

The total amount of different samples was 309, since there were 103 animals with three different ageing periods: 1, 7, and 21 days. Then, the opinions of our consumers could only be estimated inducing a preference or ranking function as explained earlier. Notice that only such functions can be used in to compare the preferences of different consumers; in general, two arbitrary consumers have not tasted samples of the same animal prepared with the same ageing. However, it is possible to compare the preference functions of any couple of consumers as vectors in a high dimension space following the kernel-based method of Section 2.3.

3.1. Clustering process

The clustering algorithm (Dubnov et al., 2002) returns the trees depicted in Fig. 1. Split nodes achieved a confidence level of 91% for the tenderness dataset, and 97% for acceptance. The leaves of these trees and the dataset of flavour reached lower confidence levels, and therefore they were rejected.

The job of clustering is to compute groups with minimal intra-group and maximal inter-group distances or differences. In our case, the relevance of clusters could be estimated, partially, by the coherence of consumers included in the same cluster, which can be measured by the classification error of the SVM used to compute the ranking or preference function of each cluster. When the algorithm joins the preference judgements of the members of the same cluster (see Eq. (12)) in order to obtain its preference function, it is very common that disagreements appear; i.e., two or more consumers in the same cluster may have a different preference about two samples. About 16% of preference pairs of each cluster express a particular disagreement with the majority opinion of the cluster (Table 4). However, every preference judgment is included in the training set of each cluster; this sums more than 2000 preference judgments, which means (see Eq. (4)) more than 4000 training set instances for the corresponding classification sets.
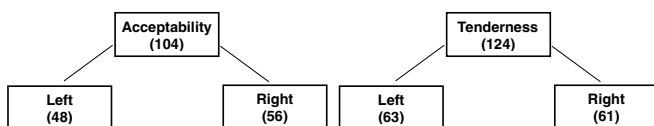
Table 4
For clusters of acceptance and tenderness datasets, this table reports the number of preference judgements (PJ), percentage of disagreements, and classification errors achieved into clusters with their own ranking or preference function, and using the function of the other cluster

Dataset cluster	PJ	Disagreements (%)	Classification errors using function	
			Own (%)	Other (%)
Acceptance				
Left	1927	16.19	19.20	50.96
Right	2150	17.07	21.12	54.95
Tenderness				
Left	2487	15.96	19.38	61.98
Right	2432	15.21	19.59	61.06

When we use a polynomial kernel of degree 2, the errors range from 19.20% to 21.12%; we used this kernel following Luaces et al. (2004), Díez et al. (2004) and Del Coz et al. (2004). Nevertheless, if we apply the induced classification function of each cluster to the other one, then the errors rise to more than 50% in the case of acceptance and more than 60% in the case of tenderness. In both cases we are ranking the same samples and these errors can be understood as the probability of reversing the order given by one such cluster when we use the criteria of the other one. Therefore, 50% of error means a random classification, and over that threshold means that the ranking criteria are approaching the exact opposite. All the details are collected in Table 4.

3.2. Analysis of market segments

It is well known that meat characteristics are mainly the result of a set of complex factors, both pre- and post-slaughter, which could be minimised through 'best practice' animal and carcass management. However, to determine genetic cause of the variation in quality extra studies are required (Maher et al., 2004). In the present study, we focus our attention on three of these: breed, slaughter weight and ageing period. Specifically, we were interested in knowing if there are different groups of people who prefer some breeds to others or who prefer long to short ageing periods or who prefer light to heavy carcasses. We used these characteristics since they are some of the most relevant features in the acceptance and tenderness appreciation of beef. Thus, Monsón et al. (2005) showed the significance of breed on meat quality, when dairy and meat breeds were compared, and Chambaz, Scheeder, Kreuzer, and Dufey (2003) showed differences in meat quality between breeds, even when animals were slaughtered with similar intramuscular fat contents. Gregory, Cundiff, Koch, Dikerman, and Koomaraie (1994) emphasised the importance of achieving a balance between carcass composition and palatability attributes for composite breeds slaughtered at optimum carcass weights. The importance of ageing in meat quality has been widely studied. Ageing improves meat texture, irrespective of other postmortem handling treatments (Jeremiah & Gibson, 2003).

Fig. 1. Trace of the clustering algorithm. In each node we report the number of consumers.

To assess the meaning of the preference criteria of each cluster, we used the ranking or preference functions to order the samples; then we assigned 10 points to those samples included in the first decile, 9 to the second decile, and so on. Average points obtained by each breed, ageing period and each pair of breed-weight are shown in Tables 5 and 6 and Fig. 2 (acceptance) and in Tables 7 and 8 and Fig. 3 (tenderness); the average score of all samples was 5.5. The results are similar if we use quartiles instead of deciles or any other division of the relative rankings of each cluster.

In the acceptance dataset (Table 5), it is possible to observe the opposite rank achieved by Retinta (RE, unimproved) and Asturiana de los Valles (AS, double muscled) breeds (see below): they were first and last (or almost last) in each cluster alternatively (Table 6). Del Coz et al. (2004) used Boolean attributes to include breed in the description of each sample, and then RE and AS were found to be the most relevant Boolean features to explain consumer acceptance of meat. Additionally, these two breeds have significant differences in carcass composition (Tables 1 and 2). AS is a double muscled breed, and therefore has the lowest percentage of subcutaneous and inter-muscular fat; while RE is an unimproved breed with the highest percentage of fat. The positive influence of fat content on acceptability
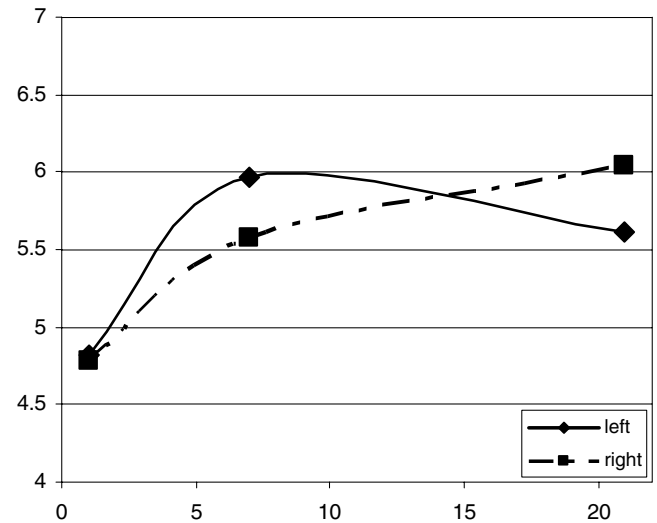

Fig. 2. Acceptance of beef meat. Average ranking scores for each ageing period.

Table 5
Average ranking scores obtained for each breed in acceptance dataset's clusters

Left			Right		
Rank	Breed	Average ranking	Rank	Breed	Average ranking
1	Retinta	6.67	1	Asturiana	6.03
2	Gallega	5.97	2	Avileña	5.73
3	Pirenaica	5.73	3	Pirenaica	5.62
4	Avileña	5.36	4	Gallega	5.47
5	Morucha	5.33	5	Parda	5.44
6	Parda	4.96	6	Retinta	5.10
7	Asturiana	4.23	7	Morucha	4.84

Table 6
For clusters of acceptance dataset, the table shows average ranking scores obtained for each pair of breed-weight

Left			Right		
Rank	Breed-weight	Average ranking	Rank	Breed-weight	Average ranking
1	Retinta_Heavy	7.29	1	Parda_Heavy	7.54
2	Gallega_Light	6.60	2	Avileña_Heavy	7.21
3	Pirenaica_Light	6.42	3	Pirenaica_Heavy	7.10
4	Retinta_Light	6.05	4	Morucha_Heavy	6.90
5	Morucha_Heavy	5.95	5	Retinta_Heavy	6.48
6	Avileña_Heavy	5.46	6	Asturiana_Heavy	6.31
7	Gallega_Heavy	5.33	7	Asturiana_Light	5.83
8	Avileña_Light	5.29	8	Gallega_Light	5.75
9	Parda_Heavy	5.04	9	Gallega_Heavy	5.22
10	Pirenaica_Heavy	4.95	10	Pirenaica_Light	4.33
11	Parda_Light	4.88	11	Avileña_Light	4.10
12	Morucha_Light	4.79	12	Retinta_Light	3.71
13	Asturiana_Light	4.79	13	Parda_Light	3.33
14	Asturiana_Heavy	3.44	14	Morucha_Light	3.08

Table 7
Average ranking scores obtained for each breed in tenderness dataset

Left			Right		
Rank	Breed	Average ranking	Rank	Breed	Average ranking
1	Pirenaica	6.44	1	Asturiana	6.55
2	Retinta	5.98	2	Gallega	6.24
3	Gallega	5.39	3	Parda	5.65
4	Avileña	5.38	4	Avileña	5.64
5	Morucha	5.09	5	Retinta	5.21
6	Asturiana	5.03	6	Pirenaica	5.02
7	Parda	4.90	7	Morucha	4.09

Table 8
For clusters of tenderness dataset, the table shows average ranking scores obtained for each pair of breed-weight

Left			Right		
Rank	Breed-weight	Average ranking	Rank	Breed-weight	Average ranking
1	Retinta_Heavy	7.86	1	Gallega_Heavy	6.89
2	Morucha_Heavy	7.19	2	Parda_Heavy	6.79
3	Pirenaica_Heavy	6.86	3	Asturiana_Light	6.67
4	Parda_Heavy	6.46	4	Asturiana_Heavy	6.38
5	Gallega_Light	6.45	5	Avileña_Heavy	6.08
6	Pirenaica_Light	6.08	6	Pirenaica_Heavy	5.71
7	Avileña_Heavy	6.04	7	Gallega_Light	5.70
8	Asturiana_Light	5.50	8	Retinta_Light	5.24
9	Avileña_Light	4.62	9	Retinta_Heavy	5.19
10	Asturiana_Heavy	4.31	10	Avileña_Light	5.19
11	Gallega_Heavy	4.22	11	Morucha_Heavy	4.57
12	Retinta_Light	4.14	12	Parda_Light	4.54
13	Parda_Light	3.38	13	Pirenaica_Light	4.42
14	Morucha_Light	3.29	14	Morucha_Light	3.67

is well recognised; even, when domestic and foreign meat are compared (Delgado et al., 2005). Therefore, fatness could be one of the reasons to assign opposite ratings to these two breeds, although, the final acceptance scorings rely on a complex set of features (Jeremiah & Gibson, 2003).
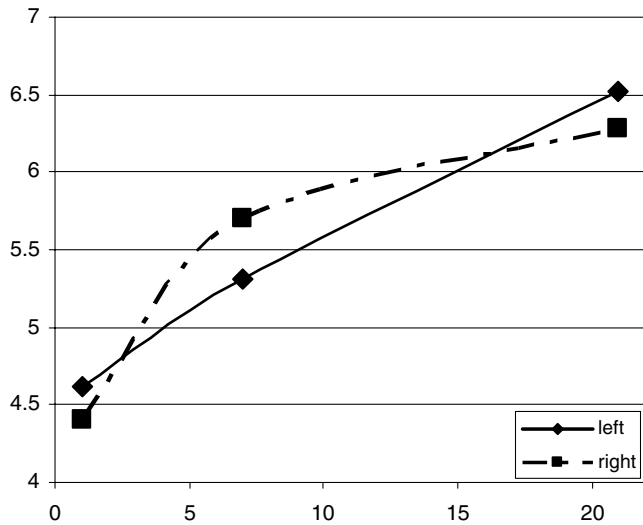
Fig. 3. Tenderness of beef meat. Average ranking scores for each ageing period.

In both clusters, Morucha (MO, unimproved breed) and Parda Alpina (PA, double purpose breed) showed the lowest ratings for acceptability, which could be related to their different fatty acid compositions or their texture behaviour. Thus, MO and PA showed higher ω-3 values ($p < 0.05$), associated with lower acceptability for Spanish people with other species (Sañudo et al., 2000), higher MUFA content ($p < 0.05$ with AS but $p > 0.05$ with the other breeds) and lower PUFA/SFA ratio ($p < 0.05$) (Insausti et al., 2004). Also, in a multi-comparison of four different cattle breeds, Monsón et al. (2005) found that PA was between two breeds with lower ratings in beef acceptance at short ageing times (1, 3, and 7 days). Also, in the same animals, the PA breed showed the highest 20% and 80% compression values in the raw meat (Sañudo et al., 2004).

On the other hand, after analysing ageing periods in both clusters (Fig. 2) we see that people in the left cluster preferred a 7 days ageing while in the right cluster longer periods have better acceptance. These results agree with those found by other authors (Bidner, Wyatt, Humes, Franke, & Blouin, 2002; Campo et al., 1999; Jeremiah & Gibson, 2003; Monsón et al., 2005), who found that desirability increased progressively during 5 weeks of postmortem ageing.

In Table 6 is shown how each group of consumers rated weight when asked to rate the *acceptance* of the samples of meat. In right cluster, samples from heavy animals are preferred to light ones for almost every breed, except Rubia Gallega. This separation is not so clear in the left cluster. In general, higher acceptability of heavier animals in reared on high-energy diets, could be related to their higher fatness level and better texture (Sañudo et al., 2004). An analysis of variance confirmed this; in the right cluster there is a significant difference between heavy and light carcasses at the 98% level. In the left cluster, no significant difference was observed.

Meat from Pirenaica and RE breeds was more tender for people in the left cluster; however, they are ranked in low positions in the right cluster (Table 7). The opposite is true of meat from AS. Again, the AS and RE breeds achieve opposite ranks in each cluster. The relationship between tenderness, considered a major factor for most consumers, and acceptability could explain the similarity of the results. As consumers usually distinguish tenderness differences very easily, one would expect better agreement between clusters. In addition, some flavour interference, as consumers are more sensitive to this parameter, could have affected the results.

Contrary to acceptability, in tenderness no marked differences between clusters' preferences regarding ageing period was seen (Fig. 3): all breeds had higher tenderness scores as ageing increased. This is not surprising, it is well documented that long ageing periods give rise to more tender meat (Maher et al., 2004; Monsón et al., 2005).

Table 8 reports ranking scores obtained for each pair of breed-weight for tenderness. Most consumers clearly prefer heavy to light animals. These results are in agreement with the instrumental values (WB) found in heated meat, which were significantly lower for samples from the heaviest animals (Sañudo et al., 2004). However, in this dataset the left cluster consumers show slight preference for heavy animals too, something that does not happen in the acceptance dataset. In both clusters, in light animals, beef breeds are preferred to unimproved ones. No such tendency could be observed in meat from the heavier carcasses. Again, an analysis of variance confirmed these observations: in the right cluster there was a significant difference between heavy and light carcasses at the 96% level while in the left cluster this only reached the 80% level.

4. Conclusions

In this paper we use artificial intelligence methods to learn the reasons why groups of consumers prefer some beef to others. It revealed the existence of different market segments characterised by divergent sensitivities in the appreciation of this kind of meat. The method stresses that it is possible to map with continuity people's preferences into a metric space, where we can compute the similarity between preference criteria.

Once we have a reasonable similarity measure for preference criteria, the main goal is to discover different groups of consumers (or market segments) and explain which kinds of products they prefer. For this purpose, we used two learning tools: (i) to group people with similar preferences, a hierarchical clustering algorithm that directly captures the intention of the similarity functions using a proximity matrix of pairwise relations; (ii) a SVM algorithm to learn preference functions using the descriptions of the products.

The method was applied to a dataset that collects ratings of a panel of consumers, and we found that there are two main groups of similar size. The most important conclusion is that they prefer opposite types of meat. One of them clearly prefers meat with the highest percentages of fat

(from unimproved rustic breeds) while the other group prefers meat with less fat. Also, one of the groups clearly prefers heavy animals to light ones. These heavier animals have, in general, more tender meat. The goal for the future of the Meat Sector and meat researchers would be to clearly identify the characteristics that define these different groups of consumers in order to provide different products to suit different customer segments.

Finally, we would like to emphasise that the approach followed in this paper has been useful for dealing with sensory data collected from panels of untrained consumers that use different scales, and that have tasted only a small portion of all available meat samples.

Acknowledgments

References

Bahamonde, A., Bayón, G. F., Díez, J., Quevedo, J. R., Luaces, O., Del Coz, J. J., et al. (2004). Feature subset selection for learning preferences: a case study. In *Proceedings of the 21st international conference on machine learning, ICML 04*, Banff, Canada (pp. 49–56).

Bidner, T. D., Wyatt, W. E., Humes, P. E., Franke, D. E., & Blouin, D. C. (2002). Influence of Brahman derivative breeds and Angus on carcass traits, physical composition, and palatability. *Journal of Animal Science, 80*, 2126–2133.

Campo, M. M., Santolaria, P., Sañudo, C., Lepetit, J., Olleta, J. L., Panea, B., et al. (2000). Assessment of breed type and ageing time effects on beef meat quality using two different texture devices. *Meat Science, 55*, 371–378.

Campo, M. M., Sañudo, C., Panea, B., Albertí, P., & Satolaria, P. (1999). Breed type and ageing time effects on sensory characteristics of beef strip loin steaks. *Meat Science, 51*, 383–390.

Chambaz, A., Scheeder, M. R. L., Kreuzer, M., & Dufey, P. A. (2003). Meat quality of Angus, Simmental, Charolais and Limousin steers compared at the same intramuscular fat content. *Meat Science, 63*, 491–500.

Cohen, W., Shapire, R., & Singer, Y. (1999). Learning to order things. *Journal of Artificial Intelligence Research, 10*, 243–270.

Del Coz, J. J., Bayón, G. F., Díez, J., Luaces, O., Bahamonde, A., & Sañudo, C. (2004). Trait selection for assessing beef meat quality using non-linear SVM. In *Proceedings of the eighteenth annual conference on neural information processing systems (NIPS 2004)*. Vancouver, British Columbia, Canada (pp. 321–328).

Delgado, E. J., Rubio, M. S., Iturbe, F. A., Méndez, R. D., Cassís, L., & Rosiles, R. (2005). Composition and quality of Mexican and imported retail beef in Mexico. *Meat Science, 69*, 465–471.

Díez, J., Bahamonde, A., Alonso, J., López, S., Del Coz, J. J., Quevedo, J., et al. (2003). Artificial intelligence techniques point out differences in classification performance between light and standard bovine carcasses. *Meat Science, 64*, 249–258.

Díez, J., Bayón, G. F., Quevedo, J. R., Del Coz, J. J., Luaces, O., & Alonso, J., et al. (2004). Discovering relevancies in very difficult regression problems: applications to sensory data analysis. In *Proceedings of the European conference on artificial intelligence (ECAI'04)*, Valencia, Spain (pp. 993–994).

Díez, J., Del Coz, J., Sañudo, C. J., Albertí, P., & Bahamonde, A. (2005). A Kernel based method for discovering market segments in beef meat. In *Proceedings of the 9th European conference on principles and practice of knowledge discovery in databases (PKDD'05)*. Porto, Portugal (pp. 462–469).

Dubnov, S., El-Yaniv, R., Gdalyahu, Y., Schneidman, E., Tishby, N., & Yona, G. (2002). A new nonparametric pairwise clustering algorithm based on iterative estimation of distance profiles. *Machine Learning, 47*, 35–61.

Gil, M., Serra, X., Gispert, M., Oliver, M., Sañudo, C., Panea, B., et al. (2001). The effect of breed-production systems on the myosin heavy chain 1, the biochemical characteristics and the colour variables of *Longissimus thoracis* from seven Spanish beef cattle breeds. *Meat Science, 58*, 181–188.

Gregory, K. E., Cundiff, L. V., Koch, R. M., Dikerman, M. E., & Koomaraie, M. (1994). Breed effects, retained heterosis, and estimates of genetic and phenotypic parameters for carcass and meat traits of beef cattle. *Journal of Animal Science, 72*, 1174–1183.

Herbrich, R., Graepel, T., & Obermayer, K. (2000). Large margin rank boundaries for ordinal regression. In A. Smola, P. Bartlett, B. Scholkopf, & D. Schuurmans (Eds.), *Advances in large margin classifiers* (pp. 115–132). Cambridge, MA: MIT Press.

Insausti, K., Berian, M. J., Alzueta, M. J., Carr, T. R., & Purroy, A. (2004). Lipid composition of the intramuscular fat of beef from Spanish cattle breeds stored under modified atmosphere. *Meat Science, 66*, 639–646.

ISO 8589. (1988).

Jeremiah, L. E., & Gibson, L. L. (2003). The effects of postmortem product handling and ageing time on beef palatability. *Food Research International, 36*, 929–941.

Joachims, T. (1999). Making large-scale SVM learning practical. In B. Schölkopf, C. Burges, & A. Smola (Eds.), *Advances in kernel methods – Support vector learning*. Cambridge, MA: MIT Press.

Joachims, T. (2002). Optimising search engines using click through data. In *Proceedings of the ACM conference on knowledge discovery and data mining (KDD)*. Edmonton, Alberta, Canada (pp. 133–142)..

Luaces, O., Bayón, G. F., Quevedo, J. R., Díez, J., Del Coz, J. J., & Bahamonde, A. (2004). Analysing sensory data using non-linear preference learning with feature subset selection. In *Proceedings of the 15th European conference of machine learning (ECML 04)*. Pisa, Italia (pp. 286–297).

Maher, S. C., Mullen, A. M., Keane, M. G., Buckley, D. J., Kerry, J. P., & Moloney, A. P. (2004). Decreasing variation in the quality of beef through homogenous pre- and post-slaughter management. *Meat Science, 67*, 33–43.

Monsón, F., Sañudo, C., & Sierra, I. (2005). Influence of breed and ageing time on the sensory meat quality and consumer acceptability in intensively reared beef. *Meat Science, 71*, 471–479.

Sañudo, C., Enser, M. E., Campo, M. M., Nute, G. R., María, G., Sierra, I., et al. (2000). Fatty acid composition and sensory characteristics of lamb carcasses from Britain and Spain. *Meat Science, 54*, 339–346.

Sañudo, C., Macie, E. S., Olleta, J. L., Villarroel, M., Panea, B., & Albertí, P. (2004). The effects of slaughter weight, breed type and ageing time on beef meat quality using two different texture devices. *Meat Science, 66*, 925–932.

Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge University Press.

Verbeke, W., & Vackier, I. (2004). Profile and effects of consumer involvement in fresh meat. *Meat Science, 67*, 159–168.

Weston, J., Elisseeff, A., Bakir, G., & Sinz, F. (2005). SPIDER: object-orientated machine learning library. Available from http://www.kyb.tuebingen.mpg.de/bs/people/spider/.