

# A Kernel Based Method for Discovering Market Segments in Beef Meat

Jorge Díez<sup>1</sup>, Juan José del Coz<sup>1</sup>, Carlos Sañudo<sup>2</sup>,  
Pere Albertí<sup>3</sup>, and Antonio Bahamonde<sup>1</sup>

<sup>1</sup> Artificial Intelligence Center, University of Oviedo at Gijón (Asturias), Spain  
{jdiez, juanjo, antonio}@aic.uniovi.es  
www.aic.uniovi.es

<sup>2</sup> Facultad de Veterinaria, University of Zaragoza, Zaragoza (Aragón), Spain  
csanudo@unizar.es

<sup>3</sup> Service of Agriculture and Food Science Research, Zaragoza (Aragón), Spain  
palberti@aragon.es

**Abstract.** In this paper we propose a method for learning the reasons why groups of consumers prefer some food products instead of others. We emphasize the role of groups given that, 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 discover significant groups of consumers with homogeneous tastes. Finally in each cluster, we learn, with a SVM, a function that explains the tastes of the consumers grouped in the cluster. To illustrate our method, a real case of consumers of beef meat was studied. The panel was formed by 171 people who rated 303 samples of meat from 101 animals with 3 different aging periods.

## 1 Introduction

Consumer 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 food products in each segment.

From a conceptual point of view, consumer panels are made up of untrained consumers; 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. Nevertheless, the Market is not interested in tastes of individual consumers, the purpose of marketing studies of sensorial data is to discover, if there 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 high dimension space; for this purpose, we present here a kernel-based method. To illustrate our method, we used a data set that

collects the ratings of a panel of beef meat consumers. The panel studied was formed by 171 people rating samples of 303 different kinds of beef meat [1] from different breeds, live weights, and aging periods.

## 2 Description of 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 kind of *batch effect*: a product will obtain a higher/lower rating when it is 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 [2]; 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 given that, in general, the size of the sample of food prevents panelists from testing all products. Hence, we cannot ask our panelists to spend long periods rating the whole set of food samples. Typically, each consumer only participates in one or a small number of testing 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 represent the descriptions of food products. Then, the similarity is defined by means of the kernel attached to the representation map.

Once we have people preferences represented in a metric space, and we have defined a similarity function, then we use a clustering algorithm. Finally, we only need to explain the meaning and implications of each cluster in the context of the food products. For this purpose, we will learn a preference or ranking function from the union of preference judgments expressed by the member of the cluster; this will provide the consensus assessment function of the cluster.

## 3 Description of the Beef Meat Experiment

To illustrate our method we used a database described in [1]. The data collects the sensory ratings of a panel of beef meat consumers about three aspects: flavour, tenderness, and acceptability.

For this experience, more than 100 animals of 7 Spanish breeds were slaughtered to obtain two kinds of carcasses: lights, from animals with a live weight around 300–350 kg (light); and heavies, from animals at 530–560 kg. The set of animals was uni-

formly distributed by breeds and weights. Additionally, to test the influence of aging in consumers' appreciation, each piece of meat was prepared with 3 aging periods, 1, 7, and 21 days. On the other hand, the 7 breeds used constitute a wide representation of beef cattle. These breeds can be divided into four types: double muscled (DM, one breed), fast growth (FG, two breeds), dual purpose (DP, one breed), and unimproved rustic type (UR, three breeds). In Table 1 for each breed, we show the average percentages of fats, muscle and bone.

**Table 1.** Carcass compositions of 7 Spanish beef breeds used in the experiment

Breed		Fat %		Bone	Muscle	Intramuscular
Name	Type	inter-muscular	subcutaneous	%	%	fat %
Asturiana Valles	DM	4.77	0.89	16.00	78.34	0.90
Avileña	UR	13.17	3.53	19.25	64.05	2.28
Morucha	UR	12.46	3.46	19.28	64.80	2.10
Parda Alpina	DP	9.65	2.32	20.86	67.17	1.82
Pirenaica	FG	9.02	3.01	17.33	70.63	1.48
Retinta	UR	14.16	4.75	20.89	60.20	2.13
Rubia Gallega	FG	5.73	1.20	16.56	76.52	1.12

Each kind of meat was also described by a panel of 11 trained experts who rate 12 traits of products such as fibrosis, flavor, odor, etc.. In this paper, we considered the average rate of each trait. The characterization of meat samples was completed with 6 physical features describing its texture.

## 4 Vectorial Representation of Preference Criteria

As was explained above, in order to compare the preference criteria of consumers we need to state a common language. We cannot use for this purpose the ratings assigned by consumers to food products, since they have rated, in general, different sets of samples. Then 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 [3, 4, 5]. Then we will try to induce a real *preference*, *ranking*, or *utility function*  $f$  from the input space of object descriptions, say  $\mathbf{R}^d$ , in such a way that it maximizes the probability of having  $f(\mathbf{v}) > f(\mathbf{u})$  whenever  $\mathbf{v}$  is preferable to  $\mathbf{u}$ ; we call such pairs, *preference judgments*. This functional approach can start from a set of objects endowed with a (usually ordinal) rating, as in regression; but essentially, we only need a collection of preference judgments.

When we have a set of ratings given by a consumer  $c$ , we must take into account the session where the ratings have been assessed [6, 7], as was explained in section 2. Thus, for each session we include in the set of preference judgments,  $PJ_c$ , the pairs  $(\mathbf{v}, \mathbf{u})$  whenever consumer  $c$  assessed to sample represented by  $\mathbf{v}$  a higher rating than to the sample represented by  $\mathbf{u}$ . In order to induce the ranking function, as in [3], we look for a function  $F_c: \mathbf{R}^d \times \mathbf{R}^d \rightarrow \mathbf{R}$  such that

$$\forall \mathbf{x}, \mathbf{y} \in \mathbf{R}^d, F_c(\mathbf{x}, \mathbf{y}) > 0 \Leftrightarrow F_c(\mathbf{x}, \mathbf{0}) > F_c(\mathbf{y}, \mathbf{0}) \tag{1}$$

Notice that the right hand side of (1) establishes an ordering of functional expressions of a generic couple  $(\mathbf{x}, \mathbf{y})$  of objects representations. This suggests the definition

$$f_c: \mathbf{R}^d \rightarrow \mathbf{R}, f_c(\mathbf{x}) = F_c(\mathbf{x}, \mathbf{0}) \tag{2}$$

The idea is then to obtain ranking functions  $f_c$  from functions like  $F_c$ , as in (2), when  $F_c$  fulfils (1). Thus, given the set of preference judgments  $PJ_c$ , we can specify  $F_c$  by means of the constraints

$$\forall (\mathbf{v}, \mathbf{u}) \in PJ_c, F_c(\mathbf{v}, \mathbf{u}) > 0 \text{ and } F_c(\mathbf{u}, \mathbf{v}) < 0 \tag{3}$$

Therefore,  $PJ_c$  gives rise to a set of binary classification training set to induce  $F_c$

$$E_c = \{(\mathbf{v}, \mathbf{u}, +1), (\mathbf{u}, \mathbf{v}, -1) : (\mathbf{v}, \mathbf{u}) \in PJ_c\} \tag{4}$$

Nevertheless, a separating function for  $E_c$  does not necessarily fulfill (1). Thus, we need an additional constraint. So, if we represent each object description  $\mathbf{x}$  in a higher dimensional feature space by means of  $\phi(\mathbf{x})$ , then we can represent pairs  $(\mathbf{x}, \mathbf{y})$  by  $\phi(\mathbf{x}) - \phi(\mathbf{y})$ . Hence, a classification SVM can induce from  $E_c$  a function of the form:

$$F_c(\mathbf{x}, \mathbf{y}) = \sum_{s \in S(c)} \alpha_s z_s \langle \phi(\mathbf{x}_s^{(1)}) - \phi(\mathbf{x}_s^{(2)}), \phi(\mathbf{x}) - \phi(\mathbf{y}) \rangle \tag{5}$$

where  $\langle \mathbf{x}, \mathbf{y} \rangle$  stands for the inner product of vectors  $\mathbf{x}$  and  $\mathbf{y}$ ;  $S(c)$  is the set of support vectors, notice that they are formed by two d-dimensional vectors  $(\mathbf{x}_s^{(1)}, \mathbf{x}_s^{(2)})$ , while the scalars  $z_s$  represent the class +1 or -1. Trivially,  $F_c$  fulfils the condition (1). Let us remark that if  $k$  is a kernel function, defined as the inner product of two objects represented in the feature space, that is,  $k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$ , then the kernel function used to induce  $F_c$  is

$$\mathbf{K}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) = k(\mathbf{x}_1, \mathbf{x}_3) - k(\mathbf{x}_1, \mathbf{x}_4) - k(\mathbf{x}_2, \mathbf{x}_3) + k(\mathbf{x}_2, \mathbf{x}_4) \tag{6}$$

Usually it is employed a linear or a simple polynomial kernel; that is,  $k(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle$ , or  $k(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle + 1)^g$ , with  $g = 2$ .

Once we have a function  $F_c$  for a consumer  $c$  fulfilling (1), then, using (2), a ranking or preference or utility function  $f_c$  is given (but for an irrelevant constant) by

$$f_c(\mathbf{x}) = \sum_{s \in S(c)} \alpha_s z_s \langle \phi(\mathbf{x}_s^{(1)}) - \phi(\mathbf{x}_s^{(2)}), \phi(\mathbf{x}) \rangle = \sum_{s \in S(c)} \alpha_s z_s (k(\mathbf{x}_s^{(1)}, \mathbf{x}) - k(\mathbf{x}_s^{(2)}, \mathbf{x})) \tag{7}$$

Therefore,  $f_c$  can be represented by the weight vector  $\mathbf{w}^c$  in the higher dimensional space of features such that

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

Notice that (8) defines the ranking of an object represented by a vector  $\mathbf{x}$ . This is not an absolute value; its importance is the relative position that gives to  $\mathbf{x}$  against to other objects  $\mathbf{y}$  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 (10) in terms of the kernels used in the previous derivations. The essential equality is:

$$\begin{aligned} \langle \mathbf{w}^c, \mathbf{w}^{c'} \rangle &= \sum_{s \in S(c)} \sum_{s' \in S(c')} \alpha_s \alpha_{s'} z_s z_{s'} \langle \phi(\mathbf{x}_s^{(1)}) - \phi(\mathbf{x}_s^{(2)}), \phi(\mathbf{x}_{s'}^{(1)}) - \phi(\mathbf{x}_{s'}^{(2)}) \rangle \\ &= \sum_{s \in S(c)} \sum_{s' \in S(c')} \alpha_s \alpha_{s'} z_s z_{s'} \mathbf{K}(\mathbf{x}_s^{(1)}, \mathbf{x}_s^{(2)}, \mathbf{x}_{s'}^{(1)}, \mathbf{x}_{s'}^{(2)}) \end{aligned} \tag{10}$$

### 5 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 [8].

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 equation (9). Then, matrix  $S$  is transformed iteratively, following a two step procedure that converges to a two values 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, in [8] the authors provide a cross validation method that measures an index that can be read as a significance level; we will only accept splits which level is above 0.90.

The first step normalizes the columns of  $S$  using the  $L_\infty$  norm; then the proximities are re-estimated using the Jensen-Shannon divergence. The idea is to formalize that two preference criteria are close (after these two steps) if they were both similar and dissimilar to analogous sets of criteria before the transformation.

### 6 Experimental Results

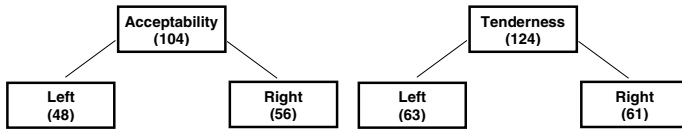
In this section, we report the outputs obtained with the database of beef meat consumers. In order to consider significant opinions, we first selected those people involved in our consumers' panel whose ratings gave rise to at least 30 preference judgments; these yielded us to consider a set of 171 panelists that tested from 9 to 14 samples of meat of 101 different animals. The total amount of different samples was 303, since the meat from each animal was prepared with 3 different aging periods: 1, 7, and 21 days. Then the opinions of our panelists can be estimated inducing a preference or ranking function as was explained in section 4. Notice that only such functions can be used in order to compare the preferences of different consumers; in general, two arbi-

**Table 2.** For clusters of acceptance and tenderness datasets, this table reports the number of preference judgments (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

trary consumers have not tested samples of the same animal prepared with the same aging. 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 4.

The clustering algorithm [8] returns the trees depicted in Figure 1. Split nodes achieved a confidence level of 91% for tenderness dataset, and 97% for acceptance. The leaves of these trees and the dataset of flavor reached lower confidence levels, and therefore they were rejected.



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

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 can be estimated, in part, by the coherence of consumers included into 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. Let us notice that the union of preference judgments of the members of the same cluster has some disagreements; if for each pair of samples we choose the most frequent relative ordering, then about 16% of preference pairs of each cluster express a particular disagreement with the majority opinion of the cluster, see Table 2. However, every preference judgment is included in the training set of each cluster; this sums more than 2000 preference judgments, what means (see equation 4 in section 4) more than 4000 training instances for the corresponding classification sets. When we use a polynomial kernel of degree 2, the errors range from 19.20% to 21.12%; we used this kernel following [2, 6, 7]. 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. Notice that 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 of such clusters when we use the criteria of the other one. Therefore, 50% of error

means a random classification, and over that threshold means that ranking criteria is approaching the exactly opposite, see Table 2.

In general, it is well known that meat qualities are mainly the result of a set of complex factors. In this study, we are interested in knowing if there are different groups of people who prefer some breeds to others. To gain insight into the meaning of the preference criteria of each cluster, we used the ranking or preference functions to order the samples of meat; then we assessed 10 points to those samples included in the first decile, 9 to the second decile, and so on. Graphical representations of the average points obtained by each breed are shown in Figure 2; notice that the average score of all samples is 5.5. The results are quite the same if we use quartiles instead of deciles or any other division of the relative rankings of each cluster.

In the acceptance dataset (Fig. 2 left), let us emphasize the opposite role played by Retinta and Asturiana breeds: they were first and last (or almost last) in each cluster alternatively. In [6, 7] we used Boolean attributes to include the breed in the description of each sample, and then Retinta and Asturiana were found to be the most relevant Boolean features in order to explain consumer’s acceptance of meat. Additionally, these two breeds have significant differences in carcass composition (see Table 1). Notice that Asturiana breed is the only double muscled breed of the sample, and then it has the lowest values in percentages of subcutaneous and inter-muscular fat, and bone; while Retinta is the unimproved rustic breed with the highest percentages of fat and bone. Therefore, there are some reasons so as to assign opposite ratings to samples of these two breeds, although, in general, the final acceptance scorings rely on a complex set of features.

In tenderness dataset (Fig. 2 right), meat from Pirenaica and Retinta breeds are the tenderest for people in left cluster, however they are ranked in low positions in right cluster. We can say exactly the opposite of meat from Asturiana and Parda breeds. Again, Asturiana and Retinta breeds play opposites roles in each cluster.

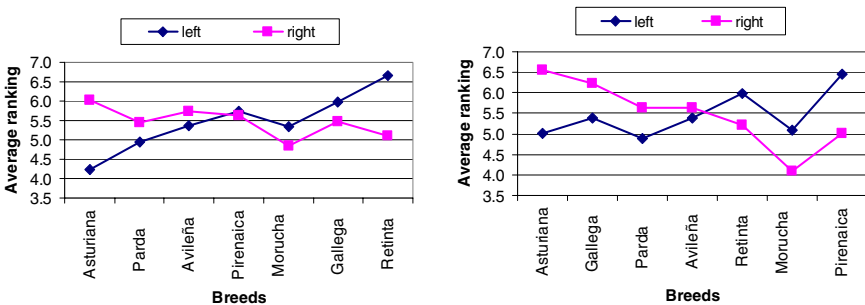


Fig. 2. Average ranking scores for each breed. Acceptance (left). Tenderness (right)

### Acknowledgments

We would like to thank: the authors of Spider [9], a MatLab toolbox that includes kernel based algorithms; Thorsten Joachims for his SVM<sup>light</sup> [10]. Those systems were used in the experiments reported in this paper; and INIA (Instituto Nacional de Inves-

tigación y Tecnología Agraria y Alimentaria of Spain) and Breeders Associations for the financial (Grant SC-97019) and technical support.

## References

1. Sañudo, C.; Macie, E.S.; Olleta, J.L.; Villarroel, M.; Panea, B.; Albertí, P.. The effects of slaughter weight, breed type and ageing time on beef meat quality using two different texture devices. *Meat Science*, 66 (2004), 925–932
2. Díez, J.; Bayón, G. F.; Quevedo, J. R.; del Coz, J. J.; Luaces, O.; Alonso, J.; Bahamonde, A.. Discovering relevancies in very difficult regression problems: applications to sensory data analysis. *Proceedings of the European Conference on Artificial Intelligence (ECAI 2004)*, 993-994
3. Herbrich, R.; Graepel, T.; and Obermayer, K.: Large margin rank boundaries for ordinal regression. In A. Smola, P. Bartlett, B. Scholkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, 115–132. MIT Press, Cambridge, MA, (2000)
4. Joachims, T.: Optimizing search engines using clickthrough data. In: *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD) (2002)*
5. Bahamonde, A.; Bayón, G. F.; Díez, J.; Quevedo, J. R.; Luaces, O.; del Coz, J. J.; Alonso, J.; Goyache, F.. Feature subset selection for learning preferences: a case study. *Proceedings of the 21st International Conference on Machine Learning, (ICML 2004)*, 49-56
6. Del Coz, J. J.; Bayón, G. F.; Díez, J.; Luaces, O.; Bahamonde, A.; Sañudo, C.. Trait selection for assessing beef meat quality using non-linear SVM. *Proceedings of the Eighteenth Annual Conference on Neural Information Processing Systems (NIPS 2004)*, 321-328
7. Luaces, O.; Bayón, G.F.; Quevedo, J.R.; Díez, J.; del Coz, J.J.; Bahamonde, A.. Analyzing sensory data using non-linear preference learning with feature subset selection. *Proceedings of the 15th European Conference of Machine Learning, (ECML 2004)*, 286-297
8. Dubnov, S.; El-Yaniv, R.; Gdalyahu, Y.; Schneidman, E.; Tishby, N.; Yona, G.. A New Nonparametric Pairwise Clustering Algorithm Based on Iterative Estimation of Distance Profiles. *Machine Learning*, 47 (2002), 35–61
9. Weston, J.; Elisseeff, A.; BakIr, G.; Sinz, F.: SPIDER: object-orientated machine learning library. <http://www.kyb.tuebingen.mpg.de/bs/people/spider/>
10. Joachims, T.. *Making large-Scale SVM Learning Practical*. *Advances in Kernel Methods - Support Vector Learning*, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, (1999)