Modeling the Organoleptic Properties of Matured Wine Distillates

S. B. Kotsiantis[1], G. E. Tsekouras [2], C. Raptis[3], P. E. Pintelas[1]

[1] Educational Software Development Laboratory
Department of Mathematics, University of Patras, Greece.
{sotos, pintelas}@ math.upatras.gr
[2] Department of Cultural Technology and Communication
University of the Aegean, Mytilene, Greece
gtsek@ ct.aegean.gr
[3] S&E&A METAXA Distilleries S. A., Athens, Greece,
craptis@metaxa.com.gr

Abstract. We present how the supervised machine learning techniques can be used to predict quality characteristics in an important chemical engineering application: the wine distillate maturation process. A number of experiments have been conducted with six regression-based algorithms, where the M5' algorithm was proved to be the most appropriate for predicting the organoleptic properties of the matured wine distillates. The rules that are exported by the algorithm are as accurate as human expert's decisions.

1 Introduction

There are many chemical engineering processes, where the quality characteristics of the product cannot be measured objectively either on-line due to the lack of proper sensors or off-line due to the absence of any measuring devices [11]. In these cases, a human expert is employed to assign the product quality characteristics to certain predefined categories (classes), based on his experience and perceptions. The procedure of employing a human expert to perform the classification usually requires the interruption of the process in order to collect a sample. Furthermore, this way of classifying the product quality is very subjective and may lead to significant errors, especially when the same expert is not always employed to perform the classification [2], [7].

A different approach to classify quality parameters is to use supervised algorithms in order to automate the process. In this paper, we present how the supervised machine learning techniques can be used to predict quality characteristics in an important chemical engineering application: the wine distillate maturation process.

The case under study is a part of the maturation process of METAXA Distilleries, a Greek aged wine distillates producing company. The firm has the problem of blending aged distillates to produce a series of final products of different quality specifications. Since most distillates characteristics cannot be accurately modelled as a function of their organoleptic properties, a system that anticipates automatically this relationship is of at most importance for the blending process engineer.

The following section describes in brief the problem and the dataset of our study. In section 3 we present the basic design issues of the supervised machine learning techniques that are used here. Section 4 compares the experimental results obtained by these techniques, while the concluding remarks are given in section 5.

2 Problem and data description

Freshly distilled spirits such as wine distillates, whiskies and rums have pungent, unpleasant odorous and sharp taste. The organoleptic properties are improved by storing the distillates in oak barrels for several years. During this process, which is known as maturation process, a number of wood components are extracted and many chemical reactions take place. However, due to the plethora of factors that affect the maturation process such as immature distillate, size, nature and usage of the barrel, environmental conditions etc., the maturation mechanisms are not completely understood. Furthermore, there is no reliable chemical or physical index that can indicate the progress of the maturation process.

The case under study is a part of the maturation process of S&E&A METAXA Distilleries S. A., a famous Greek aged wine distillates producing company. Until today, an expert tries some samples from the barrels and carries out the product quality classification, based on his perceptions. It is clear that a system, which can automatically anticipate the organoleptic properties of the distillates, based on some other accurately measured distillate characteristics, is of main importance. We applied supervised machine learning algorithms to develop models for the classification of the aroma and taste of the distillate. The intensity of aroma is of relative importance since it is basically a measure of quantity. Persistence of aroma is an indication of quality, particularly in the lingering bouquet of a mature wine.

For the aroma and taste prediction, we are based on the following input parameters:

- x1=barrel usage (the number of refills of each barrel)
- x2=barrel age (in years)
- x3=distillate age (in years)

The available data consisted of 170 input–output pairs [11]. The output values in the data set are the classifications for the aroma (y1) and taste (y2), which were given by the expert using discrete values ranging from 0 to 10, with a step of 1, where 0 and 10 correspond to the worst and finest quality, respectively. This dataset reflects long years of knowledge and experience about the process and consequently, until now it is usually used for the suggestion of suitable distillates to obtain a consistent blend from one production batch to the next.

Given ordered classes, one is not only interested in maximizing the classification accuracy, but also in minimizing the distances between the actual and the predicted classes. The usage of regression algorithms to solve ordinal classification problems has been examined in [5]. In this case each class needs to be mapped to a numeric value. Another approach is to reduce the multi-class ordinal classification problem to a set of binary classification problems using the one-against-all approach [4]. Because

the problem can be solved either with regression techniques or ordinal classification techniques we present both techniques in the next section.

3 Supervised machine learning techniques

The problem of regression consists in obtaining a functional model that relates the value of a target continuous variable y with the values of variables $x_1, x_2, ..., x_n$ (the predictors). This model is obtained using samples of the unknown regression function. These samples describe different mappings between the predictor and the target variables.

For the propose of our comparison the six most common regression techniques namely Model Trees and Rules [12], Neural Networks [6], Linear regression [3], Locally weighted linear regression [1] and Support Vector Machines [10] are used. In the following we will briefly describe these regression techniques.

Linear regression (LR) is the simplest statistical technique used to find the best-fitting linear relationship between the class and its predictors (other features).

$$y = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

Find values of beta that minimize Q:

$$Q = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}))^2$$

Note that nominal features with n values are converted into n-1 binary features and a Wald test is used to test the statistical significance of each coefficient (β_i) in the model [3].

Model trees are the counterpart of decision trees for regression tasks. Model trees are trees that classify instances by sorting them based on attribute values. Instances are classified starting at the root node and sorting them based on their attribute values. The most well known model tree inducer is the M5´ [12]. A model tree is generated in two stages. The first builds an ordinary decision tree, using as splitting criterion the maximization of the intra-subset variation of the target value [13]. The second prunes this tree back by replacing subtrees with linear regression functions wherever this seems appropriate.

M5rules algorithm produces propositional regression rules in IF-THEN rule format using routines for generating a decision list from M5´Model trees [13]. The algorithm is able to deal with both continuous and nominal variables, and obtains a piecewise linear model of the data.

Artificial Neural Networks (ANNs) are another method of inductive learning based on computational models of biological neurons and networks of neurons as found in the central nervous system of humans [6]. Regression with a neural network takes place in two distinct phases. First, the network is trained on a set of paired data to determine the input-output mapping. The weights of the connections between neurons are then fixed and the network is used to predict the numerical class values of a new set of data. Back Propagation (BP) is the most well known technique for training ANNs.

Locally weighted linear regression (LWR) is a combination of instance-based learning and linear regression [1]. Instead of performing a linear regression on the full, unweighted dataset, it performs a weighted linear regression, weighting the training instances according to their distance to the test instance at hand. This means that a linear regression has to be performed for each new test instance, which makes the method computationally quite expensive. However, it also makes it highly flexible, and enables it to approximate non-linear target functions.

The sequential minimal optimization algorithm (SMO) has been shown to be an effective method for training support vector machines (SVMs) on classification tasks defined on sparse data sets [9]. SMO differs from most SVM algorithms in that it does not require a quadratic programming solver. In [10] SMO is generalized so that it can handle regression problems (SMOreg). This implementation globally replaces all missing values and transforms nominal attributes into binary ones.

As we have previously mentioned the presented problem can be also solved by ordinal classification techniques. The most sophisticated approach that enables standard classification algorithms to make use of ordering information in ordinal class attributes is presented in [4]. This method converts the original ordinal class problem into a series of binary class problems that encode the ordering of the original classes. However, to predict the class value of an unseen instance this algorithm needs to estimate the probabilities of the k original ordinal classes using our $k - 1$ models. For example, for a three class ordinal problem, estimation of the probability for the first ordinal class value depends on a single classifier: Pr(*Target* < *first value*) as well as for the last ordinal class: Pr(*Target* > *second value*). Whereas, for class value in the middle of the range, the probability depends on a pair of classifiers and is given by

Pr(*Target* > *first value*) * (1 − Pr(*Target* > *second value*)).

4 Experiments Results

All accuracy estimates were obtained by averaging the results from 10 separate runs of stratified 10-fold cross-validation. In cross-validation technique, the training set is divided into mutually exclusive and equal-sized subsets and for each subset the regressor is trained on the union of all the other subsets. An estimation of the regressor's criterion is then the average of the error rate of each subset.

It must be mentioned that we mainly used the free available source code for our experiments by the book [13]. For our problem the regression criteria are most suitable. However, there isn't only one regressor's criterion. Table 1 represents the most well known. Fortunately, it turns out that in most practical situations the best regression method is still the best no matter which error measure is used.

Table 1. Regressors' criteria (p : predicted values, a : actual values, $\bar{a} = \frac{1}{n}\sum_i a_i$)

Mean absolute error	$\left(\left\|p_1 - a_1\right\| + \ldots + \left\|p_n - a_n\right\|\right)/n$
Root mean squared error	$\sqrt{(p_1 - a_1)^2 + \ldots + (p_1 - a_1)^2}\,/n$
Relative absolute error	$\left(\left\|p_1 - a_1\right\| + \ldots + \left\|p_n - a_n\right\|\right)\Big/\left(\left\|a_1 - \bar{a}\right\| + \ldots + \left\|a_n - \bar{a}\right\|\right)$
Root relative squared error	$\sqrt{(p_1 - a_1)^2 + \ldots + (p_1 - a_1)^2 \,/\, (a_1 - \bar{a})^2 + \ldots + (a_1 - \bar{a})^2}$

In Table 2, the regressors' criteria for each algorithm for aroma estimation are presented.

Table 2. Aroma

	M5´	*BP*	*LR*	*LWR*	*SMOreg*	*M5rules*	*Ordinal technique*
Mean absolute error	0.45	0.83	0.78	0.72	0.76	0.46	0.50
Root mean squared error	0.56	1.01	0.93	0.86	0.97	0.56	0.67
Relative absolute error	22.48	40.93	38.52	35.71	37.21	22.69	24.70
Root relative squared error	22.16	40.27	36.86	34.11	37.95	22.34	26.71

In Table 3, the regressors' criteria for each algorithm for taste estimation are presented.

Table 3. Taste

	M5´	*BP*	*LR*	*LWR*	*SMOreg*	*M5rules*	*Ordinal technique*
Mean absolute error	0.62	0.88	0.76	0.75	0.75	0.64	0.80
Root mean squared error	0.79	1.10	0.96	0.95	0.98	0.81	0.98
Relative absolute error	25.31	36.09	31.23	30.85	31.02	26.29	32.64
Root relative squared error	28.06	39.37	34.42	33.75	35.02	29.08	34.72

According to the results, the M5' is the most accurate algorithm to be used for our problem. An advantage of M5' except for its better performance is its comprehensibility. In figure 1, we present the produced rules for the prediction of wine aroma by the M5' algorithm.

```
distillateAge <= 3.33 : LM1 (51/23.386%)
distillateAge >  3.33 :
|  distillateAge <= 8 : LM2 (86/25.419%)
|  distillateAge >  8 : LM3 (33/20.786%)

LM num: 1
FinalScore = -0.1313 * barrelUsage - 0.0594 * barrelAge + 1.6481 * distillateAge + 2.0491

LM num: 2
FinalScore =     -0.2139  *  barrelUsage  -  0.0479  *  barrelAge  +0.4957*distillateAge  +
6.2112

LM num: 3
FinalScore =     -0.1134  *  barrelUsage  -  0.0705  *  barrelAge  +0.5631  *  distilla-
teAge+6.1581
```

Fig. 1. M5' model tree for aroma prediction

In figure 2, we present the produced rules by the M5' algorithm for the prediction of wine taste.

```
distillateAge <= 6.67 :
|  distillateAge <= 1.68 : LM1 (27/21.7%)
|  distillateAge >  1.68 :
|  |  barrelUsage <= 2.5 : LM2 (22/38.291%)
|  |  barrelUsage >  2.5 : LM3 (64/28.848%)
distillateAge >  6.67 : LM4 (57/21.572%)

LM num: 1
FinalScore = -0.1829 * barrelUsage - 0.1107 * barrelAge + 1.3845 * distillateAge + 2.7629

LM num: 2
FinalScore = -0.1837 * barrelUsage - 0.1958 * barrelAge + 0.9566 * distillateAge + 5.1988

LM num: 3
FinalScore =     -0.2414  *  barrelUsage  -  0.1428  *  barrelAge  +0.6525*distillateAge  +
4.8879

LM num: 4
FinalScore =     -0.1417  *  barrelUsage  -  0.1032  *  barrelAge  +1.0884*distillateAge  +
1.7945
```

Fig. 2. M5' model tree for taste prediction

It must be mentioned that the exported rules by the algorithm are as accurate as human experts' decisions.

6 Conclusion

Food processing is most often characterized by severe complexity, non-linearity and lack of objective information regarding the qualitative final product characteristics. The increasing and strong need for total quality management in food industries has rendered the construction of flexible and robust automotive decision making systems for product evaluation.

It is long been recognized that the classification of aged wine distillates is a non-linear, multi-criteria decision making problem characterized by great complexity, non-linearity and lack of objective information regarding the desired final product qualitative characteristics. The most efficient solution for the evaluation of aged wine distillates estimations with emphasis on aroma and taste, when an appropriate mathematical model cannot be incorporated, is to develop adequate and reliable expert systems based on machine learning for the classification.

In this paper, we presented how the supervised machine learning techniques can be used to predict the quality characteristics of matured wine distillates. Six algorithms were applied and compared each other with respect to real-life data, taken from a wine distillates producing company. The results showed that the M5' algorithm was the most appropriate, among the tested algorithms, for predicting the organoleptic properties of the distillates. The rules that are exported by the algorithm are as accurate as human experts' decisions.

Using machine learning in wine industry has clear advantages: it does not color the score of an individual wine with a tester's bias and it retains a level of objectivity that allows comparisons across all available wines. In a future work we will use supervised machine learning techniques to classify the variety of the wine as well as the production place (origin denomination). This classification can be carried out by processing information corresponding to physical features (color, density, conductivity, etc.) and chemical features (phenols, anthocians, amino acids, etc) [8].

References

1. Atkeson, C. G., Moore, A.W., & Schaal, S., Locally weighted learning. Artificial Intelligence Review, 11, (1997) 11–73.
2. Bende, M., & Nordin, S., Perceptual learning in olfaction: Professional wine tasters versus controls. Physiology & Behavior, 62, (1997) 1065–1070.
3. Fox, J., Applied Regression Analysis, Linear Models, and Related Methods, ISBN: 080394540X, Sage Pubns (1997).
4. Frank, E. and Hall M.: A simple approach to ordinal prediction, L. De Raedt and P. Flach (Eds.): ECML 2001, LNAI 2167, pp. 145-156, (2001), Springer-Verlag Berlin.
5. Herbrich R., Graepel T., and Obermayer K.: Regression models for ordinal data: A machine learning approach. Technical report, TU Berlin, (1999).
6. Mitchell, T., Machine Learning. McGraw Hill (1997).
7. Parr, W. V., Heatherbell, D., & White, K. G., Demystifying wine expertise: Olfactory threshold, perceptual skill and semantic memory in expert and novice wine judges. Chemical Senses, 27, (2002) 747–755.

8. Pena-Neira, A. I., Hernandez, T., Garcıa-Vallejo, C., Estrella, I., & Suarez, J., A survey of phenolic compounds in Spanish wines of different geographical origins. European Food Research and Technology, 210, (2000) 445–448.
9. Platt, J., Using sparseness and analytic QP to speed training of support vector machines. In: Kearns, M. S., Solla, S. A. & Cohn D. A. (Eds.), Advances in neural information processing systems 11. MA: MIT Press (1999).
10. Shevade, S., Keerthi, S., Bhattacharyya C., and Murthy, K., Improvements to the SMO algorithm for SVM regression. IEEE Transaction on Neural Networks, 11(5), (2000) 1188-1183.
11. Tsekouras, G., Sarimveis, H., Raptis, C., Bafas, G., A fuzzy logic approach for the classification of product qualitative characteristics, Computers and Chemical Engineering 26 (2002) 429–438.
12. Wang, Y. & Witten, I. H., Induction of model trees for predicting continuous classes, In Proc. of the Poster Papers of the European Conference on ML, Prague (pp. 128–137). Prague: University of Economics, Faculty of Informatics and Statistics (1997).
13. Witten, I.H., Frank, E., Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, San Mateo, CA, (2000).