Evaluation of Wines Rating Based on Sensory Characteristics Using Neural Networks

Nikolas Tsakiris¹, Theodoros Manavis¹ and Argyro Bekatorou²*

¹Department of Computer Science, Independent Studies of Science and Technology - University of Hertfordshire, 72 Pireos Str., 18346, Athens, Greece
²Department of Chemistry, University of Patras, 26500, Patras, Greece

Abstract: Wine is an agricultural product with very high commerce price variation, which is highly affected by quality ratings. Therefore, quality rating is particularly important for both industry and consumers. However, absence of clear concepts on what constitutes wine quality makes the perception of quality highly subjective, and it is usual for tasters to disagree on the quality rating of a specific wine. For this purpose, a Feedforward Neural Network (FNN) could be trained in order to predict wine quality. In this study, a new FNN method was developed to predict the accurate wine quality based on major sensory characteristics as FNN inputs, and to improve the ability of a taster, groups of tasters, or consumers, to rate wine by taking into account previous decisions. Specifically, five principle sensory characteristics of wines were used as inputs (Aging in Barrel, Aroma Intensity, Body, Astringency, and Acidity) in a rating range 1-3. As outputs, the quality ratings of wines in a range 70-100 were considered. The FNN was created in MATLAB with 1 hidden layer, 5 neurons and 1 output layer. For ratings divided in 5 categories the accuracy was 53% with the use of the FNN, as opposed to the accuracy of 36% achieved by Multiple Linear Regression. For ratings divided in 9 categories the accuracy was 90%. This method may allow each individual or group of tasters to introduce their own data to produce a more objective rating by taking into account previous decisions (subjective) that have accumulated in the database.

Keywords: Wine rating, sensory evaluation, Feedforward Neural Networks, regression analysis.

INTRODUCTION

According to the World Health Organisation, the world consumption in 2010 was 6.2 litres of pure alcohol per person aged 15 or older; 8% of which was consumed in the form of wine (Figure 1) [1,2]. In 2015, the world wine production according to the International Organisation of Vine and Wine is estimated at 275.7 mhl, which is slightly higher (2%) compared to 2014 and considered a good average [3]. Although, wine is an alcoholic beverage, a plethora of beneficial health effects have been attributed to moderate wine (mainly red) consumption, due to the presence of considerable amounts of bioactive phytochemicals, such as polyphenols (flavonoids, including anthocyanins, stilbenes, and proanthocyanidins) [4]. Despite these well established beneficial effects, consumer preference and price variations highly depend on the organoleptic quality of wine. Therefore, quality is highly important to the strong global wine market, but at the same time hard to define precisely. Some suggest that the absence of faults is a crucial determinant of quality, which would be accepted as at least a precursor to quality, if not part of the core of quality itself [5]. Others claim that wine must be “fit for its purpose” [6]. The difficulty faced by the wine industry in establishing quality has produced varying approaches for its evaluation during wine tasting. Influential oenologists focused on pleasure, suggesting it is the sole fundamental indicator of how good a wine is [7], or considered more quality indicators in the make-up of wine such as complexity, balance, personality, length, intensity of flavour, and varietal purity [8,9]. The nature of wine quality, with its quasi-aesthetic character and relationship to personal taste, is difficult to understand and identify [10,11] even by trained panels where subjectivity is still an unsolved problem. The perception of wine quality by consumers also depends on many factors such as reputation, “appellation d’ origine”, price, advertising, age, harvest, alcohol content, variety, taste, aroma, and colour [12-14]. Some studies have examined the components of wine quality focusing mainly on quality in relationship to purchase rather than the consumption experience (perceived rather than actual quality) [10,15]. Thus the issue of whether or not quality exists or how it functions was sidelined from the consumer’s perspective. Zeithaml [10], having distinguished perceived from objective quality, suggested that the former was a “higher level abstraction” rather than a product attribute, therefore, it is a comparative rather than a discrete assessment. Holbrook and Corfman [16], suggested that perceived quality is a “global value judgment”, meaning that quality is a broad “overview” and not a precise and partitioned concept.

Wine quality may be evaluated by blind comparative sensory tests from well-trained tasters based on
predetermined criteria and further statistical analysis [17,18]. The blind estimation depends only on sensory evaluation and not on other factors. Several statistical methods exist for the analysis of consumer research data. The most widely used are Multiple Linear Regressions (MLR), Principal Component Regression (PCR), and Partial Least Squares Regression (PLS), which rely on the linear nature of the variables. These models can take account of non-linearity by introducing transformation of the X-variables increasing the computational burden and making hardware implementation of the algorithm more complex. Despite these limitations, regression methods are widely used.

Sensory perceptions are closely connected with complicated brain functions and contents, such as memory, culture, values or emotions, which bring together knowledge or memory of a food and sensory reactions to it. Thus, an integrated perception is created that determines the ideas and emotions inevitably associated with a specific food [19]. This first approach in terminology, with words such as perception, connection, and brain shows the connection between quality and neural networks. An Artificial Neural Network (ANN) is formed by different layers: an initial layer of neurons, where the data of the input values are collected, one or more layers of intermediate neurons, where the learning processes take place, and a last one of output neurons, in which the value sought is obtained. In the training phase, the ANN learns the relations between inputs and outputs by providing a set of example training pairs and an algorithm that compares the network outputs with the expected targets. Recently, nonlinear and complex machine learning approaches such as ANNs were shown to be more reliable than conventional statistical approaches and the best alternatives to overcome their limitations due to their robustness to noise and inconsistencies in the data [20-22]. ANNs have solved a large amount of complex problems related to different disciplines (classification, clustering, regression, etc.) and are widely used in many areas such as classification, pattern recognition, etc. [23].

A multilayer feed-forward ANN has been found to be more suitable for overcoming prediction related problems particularly when a non-linear relationship exists between input and output variables [23,24]. ANNs have been successfully applied for predicting food quality [25-29], including wine [29-31]. Frances and Gordon [32] used ANNs to classify Scotch whisky samples. The ANN model was used to classify wine samples in six different regions based on the measurements of trace amounts of different metals [30]. Cortez et al. [31] predicted human wine taste preferences by their physicochemical properties. In this study, the main objectives were to predict the accurate wine quality based on major sensory characteristic inputs, and to improve the ability of a taster, groups of tasters, or consumers, to rate wine by taking into account previous decisions.
METHODS

FNN Development

A Feedforward Neural Network (FNN) was created using the Neural Network Toolbox™ of MATLAB™ (release 2008b) to evaluate the rating of a wine. In order to achieve this, a database was used, containing 9 characteristics of wines at a range of 1-3 according to their intensity. Several combinations of numbers of layers and neurons were made in order to create the best possible architecture. Also the rating was divided in 5 categories and was transformed in binary form. By means of a statistical analysis in SPSS, from the above 9 characteristics, those that are significant were identified. Confusion matrices were created by using adequate MATLAB functions (“newpr”) to visualize the performance of the FNN. MMRE (Mean Magnitude of Relative Error) and PRED 25 (Percentage Relative Error Deviation) were used to find the generalization ability of the FNN.

Original Data and 9 Sensory Characteristics of Wines

Data were obtained from the Greek wine guide “A Wine Guide 2011” [33] that is being published continuously during the last 10 years and contains data from more than 1000 Greek wines. In this guide, the writers provide a string text of about 20-50 words describing every wine regarding colour, aroma and taste, the rating of wines between 70-94 or between 1-5 (70-74=1, 75-79=2, 80-84=3, 85-89 =4, and 90-94=5), and 9 principal sensory characteristics: Colour, Sweetness, Carbon Dioxide, Aging in a Barrel, Aroma Intensity, Aging, Body, Astringency and Acidity. According to their sensory description every characteristic is evaluated at a range 1-3 according to intensity (Table 1).

Training of the FNN

A file was created in order to contain some of the original data. The ratings and 9 principal characteristics, namely Colour (code 1), Sweetness (code 2), Carbon Dioxide (code 3), Aging in a Barrel (code 4), Aroma Intensity (code 5), Aging (code 6), Body (code 7), Astringency (code 8), and Acidity (code 9), were used as inputs to the FNN. Sensory description for every characteristic was estimated at a range 1-3 according to intensity and a code (mark) indicating the kind of every characteristic. The original data contained a total of 27 different gradients.

In MATLAB, the 9 characteristics were used as input and the rating of wines as the output. Nine hundred wines were used as training data (indices from 101 to 1,000) and 100 wines were used as test data (indices 1 to 100).

After several tests, by changing the proportion between inputs and target, the form of rating from numeric to binary numeral, a FNN with 1 hidden layer with 5 neurons was found to be adequate to handle the proposed problem with 1 hidden layer, with 5 neurons and 1 output layer. Training of the FNN consisted of finding the weights that minimized error. The transfer function of the hidden layer was a tangential sigmoid one (since the problem is non-linear) and the transfer function of the output layer was linear (the combination of sigmoid and linear transfer functions is very common for this type of problems).

For the rating of wines between 70-94, a criterion between outputs and targets for training convergence was chosen with an error tolerance of 2, and for a rating between 1-5, the values of outputs that are equal to the target value were considered to be correct.

Testing the Performance of the FNN

The most widely used evaluation criterion, to assess the performance of software prediction models, is the Mean Magnitude of Relative Error (MMRE) [34]. Calculation of the MMRE is based on the absolute difference between target data and output data as % of target data. MMRE is given as the average of all the differences between inputs and target, the form of rating from numeric to binary numeral, a FNN with 1 hidden layer with 5 neurons was found to be adequate to handle the proposed problem with 1 hidden layer, with 5 neurons and 1 output layer. Training of the FNN consisted of finding the weights that minimized error. The transfer function of the hidden layer was a tangential sigmoid one (since the problem is non-linear) and the transfer function of the output layer was linear (the combination of sigmoid and linear transfer functions is very common for this type of problems).

For the rating of wines between 70-94, a criterion between outputs and targets for training convergence was chosen with an error tolerance of 2, and for a rating between 1-5, the values of outputs that are equal to the target value were considered to be correct.

<table>
<thead>
<tr>
<th>Wine number</th>
<th>Rating 70-94</th>
<th>Rating 1-5</th>
<th>Colour</th>
<th>Sweetness</th>
<th>Carbon Dioxide</th>
<th>Barrel</th>
<th>Intensity</th>
<th>Body</th>
<th>Astringency</th>
<th>Maturity</th>
<th>Acidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>996</td>
<td>76</td>
<td>2</td>
<td>11</td>
<td>21</td>
<td>31</td>
<td>41</td>
<td>52</td>
<td>62</td>
<td>70</td>
<td>82</td>
<td>92</td>
</tr>
<tr>
<td>997</td>
<td>74</td>
<td>1</td>
<td>11</td>
<td>21</td>
<td>31</td>
<td>41</td>
<td>52</td>
<td>62</td>
<td>70</td>
<td>82</td>
<td>92</td>
</tr>
<tr>
<td>998</td>
<td>88</td>
<td>4</td>
<td>13</td>
<td>21</td>
<td>31</td>
<td>43</td>
<td>53</td>
<td>63</td>
<td>72</td>
<td>82</td>
<td>92</td>
</tr>
<tr>
<td>999</td>
<td>87</td>
<td>4</td>
<td>11</td>
<td>21</td>
<td>31</td>
<td>41</td>
<td>52</td>
<td>63</td>
<td>70</td>
<td>81</td>
<td>92</td>
</tr>
<tr>
<td>1000</td>
<td>80</td>
<td>3</td>
<td>12</td>
<td>21</td>
<td>31</td>
<td>41</td>
<td>52</td>
<td>62</td>
<td>70</td>
<td>82</td>
<td>92</td>
</tr>
</tbody>
</table>
relative errors described above. PRED 25 is similarly based on the same relative errors, namely it is calculated as the percentage of all the relative errors smaller than 0.25 (a value that is considered to correspond to an acceptable prediction model). Contrary to MMRE, high PRED values are desirable. More specifically, values of PRED higher than 0.75 are considered to be good, whereas MMRE values smaller than approximately 0.2 are acceptable.

**Significance and Principal Component Analysis of the 9 Sensory Characteristics**

Using SPSS, from the 9 characteristics of wine, the data that are significant were identified. Rating was selected as the dependent variable and the 9 sensory characteristics (Colour, Sweetness, Carbon Dioxide, Barrel, Aroma, Body, Astringency, Aging, and Acidity) were selected as independent. Also in SPSS, a Principal Component Analysis (PCA) was performed for the 9 sensory characteristics in order to observe the correlation between them.

**Creation of the Confusion Matrix**

A confusion matrix is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning like artificial intelligence. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. Values in the diagonal represent a correct prediction of the FNN, while values in the other cells show failure of the FNN model to give correct results. For this purpose, using MATLAB, the function “newpr” was created, which creates a pattern recognition network and returns a FNN exactly as the “newff” does, but with an output layer transfer function “tansig” and additional plotting functions.

**RESULTS AND DISCUSSION**

**The FNN**

The purpose of this work was to predict the rating of a wine based on the sensory characteristics description. A FNN with 5 hidden layers and 1 output layer was created in MATLAB. Nine principal characteristics of wines (Colour, Sweetness, Carbon Dioxide, Aging in Barrel, Aroma Intensity, Aging, Body, Astringency, and Acidity) were used as input, at a range of 1-3 and a rating between 70-94 as target. A total of 900 wines were used as data for training and 100 for testing. Values that were 2 rating points higher or lower than the target value were considered to be correct. As a result, by the FNN a success rate of 53% was achieved. The respective result from the MLR analysis in SPSS for the same number of testing wines (100) was only 36%. This is a good indication that a FNN may be a better candidate for solving the specific problem. The R-square value indicates that about 0.56% of the variance in rating is explained by the 9 predictor variables.

For correct results, values that are 5 rating points above or under the target have a 90% success. As expected, increasing the interval that “corresponds” to success case increases the success rate.

The figure of the performance of the FNN, as produced by MATLAB, did not indicate any major problems with the training. The validation and test curves were very similar. The regression plot showed the relationship between the outputs of the FNN and the targets. If the training was perfect, the FNN outputs and the targets would be exactly equal, but the relationship is rarely perfect. In practice the R=0.72805 value shows a sufficiently good relationship between outputs and targets. Similar were the results from SPSS, where regression was calculated to be R=0.748 and this indicates that the model fits the data well.

As mentioned in previous sections, the relative errors MMRE and PRED, which indicate the prediction quality of a model, were calculated to be 0.0345 (a very small value) and 100%, respectively. This result is a good indication that the FNN model performs well in this problem.

**Changes in Order to Improve Performance of the FNN**

Apart from the target rating 70-94, were a success rate of 53% was achieved, for the target rating 1-5 a success rate of 52% was achieved, which is very similar. Subsequently the rating 70-94 was divided in 3 categories. As was expected, this improved the results and a success rate of 77% was achieved.

In order to improve the results, 700 data for training and 300 for testing were used. The result (success rate 52%) was similar. Usually in problems such as the classic “Iris problem”, much less data (150) exist, so there should be a sufficient number of both inputs and outputs. In our case, having a much larger number of data (1000) means that there were enough for both training and testing. Therefore, apparently the relative proportion (10% or 30%) did not affect the effectiveness of the method.
The target rating (70-94) was transformed into 5 categories using a binary representation. This categorization is almost the same as the precedent (2 rating points above or under the target value). In this case the results were worse (success rate 33%). The target rating was transformed in binary form in 3 categories. Then a success rate of 70% was achieved.

**MLR in SPSS and Significance of 9 Wine Characteristics**

From the 9 wine characteristics (data entry) the data that are significant (sig<0.05) were estimated with the aid of SPSS 17. Among them the following 5 characteristics seemed important: Aging in Barrel (sig=0.001), Aroma Intensity (sig=0.000), Body (sig=0.000), Astringency (sig=0.004), and Acidity (sig=0.002). The results are plausible, since for example, Colour, Sweetness, and Carbon Dioxide affect the type of wine but not how good it is. Colour (sig=0.608), Sweetness (sig<0.861), Carbon Dioxide (sig<0.550), and Aging (sig<0.352) don’t seem to be important. Aging is not always positive or negative on wine quality since it can either improve or degrade quality.

The multiple R shows a substantial correlation among the 9 predictor variables and the dependent variable Rating (R=0.748). The R-square value indicates that about 0.56% of the variance in Rating is explained by the 9 predictor variables. B gives the coefficients and constants for the MLR equation (B coefficient constant= -356.763; B coefficient for Colour= -0.095, for Sweetness = +0.028, for Carbon Dioxide= +0.160, for Aging in Barrel= +0.232, for Aroma= +2.792, for Body= +4.061, for Astringency= - 0.431, for Aging= 0.137, and for Acidity= 0.493):

\[
\text{Rating} = -356.763 - 0.095 \times \text{Colour} + 0.028 \times \text{Sweetness} + 0.160 \times \text{Carbon Dioxide} + 0.232 \times \text{Aging in Barrel} + 2.792 \times \text{Aroma} + 4.061 \times \text{Body} - 0.431 \times \text{Tannins} + 0.137 \times \text{Aging} + 0.493 \times \text{Acidity}
\]

For example, for the wine number 996, in Table 1:

\[
\text{Rating} = -356.763 - 0.095 \times 11 + 0.028 \times 21 + 0.160 \times 31 + 0.232 \times 41 + 2.792 \times 52 + 4.061 \times 62 - 0.431 \times 70 + 0.137 \times 82 + 0.493 \times 92 = 80.638.
\]

A new FNN with these 5 characteristics, 900 data for training and 100 for testing, was created. Values that were 2 rating points higher or lower than the target value were considered to be correct. Then as a result, by the FNN a success rate of 50% was achieved, which is almost the same as with the one acquired from the 9 characteristics (53%). This means that it is not necessary to dominate characteristics.

**PCA of 9 Wine Characteristics**

PCA was performed using SPSS and the results are shown in Figure 1. A correlation of axes in aroma and body can be observed, which is expected since in a wine with rich body habitually has corresponding rich aroma. Also the wines from barrel have increased
sensation of aging. It can also be observed that acidity and carbon dioxide appear with high correlation since carbon dioxide contributes to increased sensation of acidity.

Confusion Matrix by MATLAB ‘newpr’ of 9 Characteristics and 1-5 Rating

Values followed a distribution (Shapiro-Wilk Test in SPCC is 0.251, greater than 0.05) and were concentrated near the middle category 3. To get a more general view of the results, a confusion matrix was created with the function “newpr” (Table 2). The total success rate was 52.5%, which is sufficiently high. It should be noted that the FNN has the ability to provide the correct value (target) and also, when it fails, it does not do it at random, as it gives more values near the right value. For example in the case of target 2 (Table 2), 25 results were correct (i.e. 2), 2 results were 1 instead of 2, 17 were 3, 2 were 4, and none was 5. If the results were completely random it would give each one of the 5 cases the probability for a random response 1/5=20% response per category. Category 2 contains 50 cases (150 in total). Successes were 25, namely 50% of the total cases, well above the amount of random. However, there seemed to be a problem in calculating the integer values in categories 1 and 5 where the data were very few. This seems to produce poor training of neural networks in extreme ranges and so neural networks do not seem to perform well in extreme ratings.

Table 2: Results of Confusion Matrix for 1-5 Rating

<table>
<thead>
<tr>
<th>Output Class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6.7%</th>
<th>50.0%</th>
<th>75.0%</th>
<th>37.0%</th>
<th>22.2%</th>
<th>52.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Confusion Matrix</td>
<td>0 (0%)</td>
<td>2 (1.0%)</td>
<td>9 (4.5%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>33.3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10 (5.0%)</td>
<td>25 (12.5%)</td>
<td>9 (4.5%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>56.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4 (2.0%)</td>
<td>20 (10.0%)</td>
<td>60 (30.0%)</td>
<td>23 (11.5%)</td>
<td>1 (0.5%)</td>
<td>55.6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0 (0%)</td>
<td>3 (1.6%)</td>
<td>9 (4.5%)</td>
<td>17 (8.5%)</td>
<td>6 (3.0%)</td>
<td>48.6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>2 (1.0%)</td>
<td>6 (3.0%)</td>
<td>2 (1.0%)</td>
<td>20.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6.7%</td>
<td>50.0%</td>
<td>75.0%</td>
<td>37.0%</td>
<td>22.2%</td>
<td>52.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target Class</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSION

The aim of this work was to evaluate wines quality based on major sensory characteristics. A FNN was used to process this data, predict wine quality and permit a more objective and automatic evaluation of wines rating. In this new proposed method, 5 principal characteristics of wines were used as input, namely Aging in Barrel, Aroma Intensity, Body, Astringency, and Acidity. A FNN was created in MATLAB with 1 hidden layer, 5 neurons and 1 output layer. For ratings divided in 5 categories the accuracy was 53% with the use of the FNN, as opposed to the accuracy of 36% achieved by MLR. For ratings divided in 9 categories the accuracy was 90%. This method may solve the problem of subjectivity in wines rating for every evaluation done by individuals such as wine journalists, consumers or trained tasters. In this case, each individual or group of tasters can introduce their own data to produce a more objective rating by taking into account previous decisions (ratings accumulate in the data base). The ultimate goal is to start from a subjective data base, in order to create an automatic and more objective scoring based on major sensory characteristics of wine.

REFERENCES


Received on 27-05-2016
Accepted on 28-08-2016
Published on 05-12-2016

DOI: http://dx.doi.org/10.6000/1927-3037.2016.05.04.3

© 2016 Tsakiris et al.; Licensee Lifescience Global. This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.