Advance Journal of Food Science and Technology 9(3): 177-182, 2015
DOI: 10.19026/ajfst.9.1987
ISSN: 2042-4868; e-ISSN: 2042-4876

Submitted: February 6, 2015 Accepted: March 1, 2015 Published: August 10, 2015

Research Article

Multivariate Statistical Analysis Applied in Wine Quality Evaluation

[1]Jieling Zou, [2]Yin Luo, [3]Hui Zou, [4]Qi Xiu and [3]Junyan Tan
[1]College of Economics and Management,
[2]College of Information and Electrical Engineering,
[3]College of Science,
[4]College of Food Science and Nutritional Engineering, China Agricultural University,
Beijing 100083, China

Abstract: This study applies multivariate statistical approaches to wine quality evaluation. With 27 red wine samples, four factors were identified out of 12 parameters by principal component analysis, explaining 89.06% of the total variance of data. As iterative weights calculated by the BP neural network revealed little difference from weights determined by information entropy method, the latter was chosen to measure the importance of indicators. Weighted cluster analysis performs well in classifying the sample group further into two sub-clusters. The second cluster of red wine samples, compared with its first, was lighter in color, tasted thinner and had fainter bouquet. Weighted TOPSIS method was used to evaluate the quality of wine in each sub-cluster. With scores obtained, each sub-cluster was divided into three grades. On the whole, the quality of lighter red wine was slightly better than the darker category. This study shows the necessity and usefulness of multivariate statistical techniques in both wine quality evaluation and parameter selection.

Keywords: BP neural network, information entropy, principal component analysis, weighted cluster analysis, weighted TOPSIS method, wine quality evaluation

INTRODUCTION

Wine is widely consumed in many countries around the world (Bentlin *et al*., 2012) and people are increasingly concerned with the quality of the wine. Some appraise wine quality by sensory tasting while others evaluate quality of wine by physicochemical analysis. Measurement of physicochemical index technology such as heterosexual natural isotopic fractionation and nuclear magnetic resonance technology, have been gradually developed (Jiang *et al*., 2008). With the improvement of measurement techniques, physicochemical analysis is being widely used.

The methods of physicochemical specifications analysis mainly include traditional statistical methods such as Principal Component Analysis (PCA), Cluster Analysis (CA), Discriminate Analysis (DA) and Decision Trees (DT), Artificial Neural Networks (ANN) and Support Vector Machine (SVM), which have been frequently used in the field of classification (Hernanz *et al*., 2007; Aly, 2005; Kavuri and Kundu, 2011; Jin, 2005; Osorio *et al*., 2008). Two principal components were grasped by using PCA and then wine samples were clearly clustered into two homogenous groups by using CA, which was sufficient to differentiate the wines produced with different clones (Burin *et al*., 2011). But previous researchers didn't take the clustering index weights into account. The quality of cluster is largely under the influence of index weights. Cluster weights reflect the importance of the index, which is the advantage of weighted cluster analysis. In addition, the new fuzzy clustering algorithm which defines indexes weights in the framework of Axiomatic Fuzzy Set (AFS) theory is based on Shannon Entropy (Zhang *et al*., 2009). With three-layer feed forward architecture, ANN of back propagation learning was applied to update weights (Shoemaker *et al*., 1991). The method of DA was used to distinguish wines from different countries based on a minimal number of the most important parameters (Römisch *et al*., 2006). ANN methods were used for the classification of Slovak white varietal wines with the aim to classify wines by different variety, producer, location and the year of production (Kruzlicova *et al*., 2009).

Technique for Order Performance by Similarity to Ideal Solution (TOPSIS), a Distance Comprehensive Evaluation Method, is one of the most common methods for problems involving multi-criteria decisions (Cruz-Ramiaírez *et al*., 2010). To achieve competitive edge in the market, TOPSIS method was performed to select fruits from superior locations in terms of total natural antioxidants of the fruit (Sun *et al*., 2011). But

Corresponding Author: Hui Zou, College of Science, China Agricultural University, Beijing 100083, China

each indicator was given the equal weight, which can't explain the degree of importance of indicators. A comprehensive evaluation model of coal mine safety, established by the entropy weights and TOPSIS, was applied to evaluate safety conditions of production in four coal mines (Li *et al*., 2011).

In this study, we used Principal Component Analysis to eliminate the correlation between indicators. And then the wine samples were clustered by Weighted Cluster Analysis, where weights were determined by information entropy. In addition, in order to verify the accuracy of the weights, we used Back Propagation (BP) Neural Network to update weights. Finally, we used weighted TOPSIS method to evaluate the quality of various types of wine and determine the grade of wine. It is worth mentioning that the weights were respectively determined by information entropy method for red wine of the first category and the second category. Likewise, BP neural network was used to test the accuracy of the weights.

MATERIALS AND METHODS

Data sources and original indicators: Research data is quoted from the 2012 China Undergraduate Mathematical Contest in Modeling, with 27 kinds of red samples monitoring 12 parameters as a case study (http://www.mcm.edu.cn/problem/2012/2012.html).

Physicochemical indicators of red wine include Anthocyanins, Tannins, total phenols, flavonoids, resveratrol, the DPPH half inhibition of volume, L* (D65), a* (D65), b* (D65), H (D65), C (D65) and aromatic.

Table 1 shows the nature of the physicochemical indicators.

Table 1: The nature of the physicochemical indicators

Physicochemical indexes	Meaning
Anthocyanins	Antioxidants
Tannins	Indicator of wine's flavor, structure and texture
Total phenols	Evaluation of antioxidant activity
Flavonoids	Antioxidants
Resveratrol	Benefit for cardiovascular disease and cancer
DPPH semi-inhibition volume	Content for the antioxidant
L* (D65)	Lightness
a* (D65)	Color antagonistic dimension (red/green)
b* (D65)	Color antagonistic dimension (yellow/blue)
H (D65)	Hue angle
C (D65)	Color saturation
Aromatic	Aroma

PCA of indicators: A widely used multivariate analytical statistical technique, Principle Component Analysis can simplify a set of dependent texture variables to a smaller set of underlying variables based on patterns of correlation among the original variables (Lawless and Heymann, 1999). PCA can use fewer new variables instead of the original variables with the largest variability (He *et al*., 2007).

Information entropy weighted clustering: Cluster Analysis is a tool of exploratory data analysis to solve classification problems. The degree of association is strong between members of the same cluster and weak between members of different clusters (Burin *et al*., 2011). Cluster quality is largely under the influence of the weights of features. Shannon Entropy is used to defines indexes weights (Zhang *et al*., 2009).

Below are steps for weighted information entropy cluster:

- Normalize the original data matrix. Let m stands for wine samples, n is located as physicochemical indicators, F_i is the score of j-th principal component in wine sample i. Let r_{ij} is the standardization of F_i. Normalized equation is as follows:

$$r_{ij} = \frac{F_{ij} - \min_i F_{ij}}{\max_i F_{ij} - \min_i F_{ij}} \tag{1}$$

Under the j-th index, value of i-th sample valuation is p_{ij}:

$$p_{ij} = \frac{r_{ij}}{\sum_{i=1}^{m} r_{ij}} \tag{2}$$

- Calculate weights of the properties. Information entropy of j-th index is:

$$e_j = -\frac{1}{\ln m}\sum_{i=1}^{m} p_{ij} \cdot \ln p_{ij} \tag{3}$$

Below is the formula of j-th index of entropy weights w_j:

$$w_j = \left(1 - e_j\right) \Big/ \sum_{j=1}^{n}\left(1 - e_j\right) \tag{4}$$

- Use weights to calculate the squared Euclidean distance
- Do clustering analysis using ward method with squared Euclidean distance
- Analyze evaluation results

We applied a BP neural network model in iterating weight calculated by entropy method for weights accuracy inspection. BP neural network, the most widely used neural network model, is a multi-layer network model of one-way communication (Xie *et al*.,

2012). Normalized data of the red wine's main constituent was regarded as input and weight determined by information entropy was regarded as output. Component weight calculated by information entropy is definitely accurate if there is little difference between iterative weights and initial weights.

Comprehensive evaluation based on TOPSIS method: TOPSIS, developed by Hwang and Yoon (1981), is a simple ranking method in conception and application (Hwang and Yoon, 1981). The standard TOPSIS method attempts to choose alternatives that simultaneously have the shortest distance from the positive ideal solution and the farthest distance from the negative-ideal solution. Making full use of attribute information, TOPSIS provides a cardinal ranking of alternatives and does not require attribute preferences to be independent (Chen and Hwang, 1992; Yoon and Hwang, 1995). The evaluation object is ranked in accordance with the value of the relative degree of approximation. The bigger the value, the better the evaluation object.

RESULTS AND DISCUSSION

Analysis of the outcome of PCA: As is shown in Table 2, a total 89.06% of data information was explained by four principal components. So it was reasonable to take the principal components F_1, F_2, F_3, F_4 to represent the original 12 targets to conduct the cluster analysis.

The matrix of the red wine component score coefficients are represented in Table 3.

From Table 3, we knew that component 1 of the red wine contained information of anthocyanin, Tannins, total phenols, Flavonoids, DPPH Semi-inhibition volume, which could be accordingly named taste factors; Component 2 of the red wine contained information of a* (D65), C (D65), which could be named chromaticity factors; Component 3 of the red wine contained information of H (D65), b* (D65), which could be named cool tone factors; Component 4 contained information of aromatic, L* (D65), resveratrol, which could be named incense factors.

Analysis of information entropy weighted cluster: We calculated the entropy weights of four principal components of the red wine. The results of our calculation are shown in Table 4. The weights of principal components will be greater if more information is contained in the main ingredient. It indicates that the principal components are very important when they have high weights. As was shown in Table 4, we knew that the principal component 1 had the greatest impact on wines clustering.

In Table 4, we can see that iterative weights calculated by the BP neural network had a small difference from weights before iterating, which proved that weights determined by information entropy had a high accuracy.

We divided samples into different categories, based on the standard that the distance between the two classes was greater than 10 and the within-class distance was about 5.

Results for the red wine classification are shown in clustering tree (Fig. 1). According to the standard, we

Table 2: Characteristic values and contribute rate of red wine principle component

Principle component	Eigen value	Contribution rate (%)	Cumulative contribution rate (%)
F_1	5.495	45.794	45.794
F_2	2.488	20.729	66.524
F_3	1.797	14.971	81.495
F_4	0.908	7.568	89.063

Table 3: Red wine ingredient scoring matrix

Physicochemical indexes	Component 1	Component 2	Component 3	Component 4
Anthocyanins	0.152	-0.090	0.132	-0.231
Tannins	0.171	0.011	-0.029	-0.107
Total phenols	0.178	0.028	-0.023	-0.013
Flavonoids	0.167	0.023	-0.009	0.065
Resveratrol	0.078	0.230	0.047	0.350
DPPH semi-inhibition volume	0.174	0.059	-0.033	0.040
L* (D65)	-0.150	-0.116	-0.074	0.363
a* (D65)	-0.065	0.335	0.200	-0.183
b* (D65)	-0.008	0.266	-0.396	-0.054
H (D65)	-0.025	-0.037	0.519	-0.229
C (D65)	-0.066	0.360	0.088	-0.192
Aromatic	0.043	0.092	0.236	0.807

Table 4: Entropy values for principle component of red wine

Red wine	Component 1	Component 2	Component 3	Component 4
Entropy	0.917	0.977	0.964	0.971
Cluster weights	0.485	0.134	0.210	0.171
Iterative weights	0.489	0.135	0.214	0.175

Table 5: TOPSIS weights for principle component of red wine

Wine	Red	Component 1	Component 2	Component 3	Component 4
First	Weights	0.2939	0.1423	0.4447	0.1192
category	Iterative weights	0.2956	0.1405	0.4490	0.1203
Second	Weights	0.3354	0.2629	0.1961	0.2056
category	Iterative weights	0.3350	0.2628	0.1983	0.2060

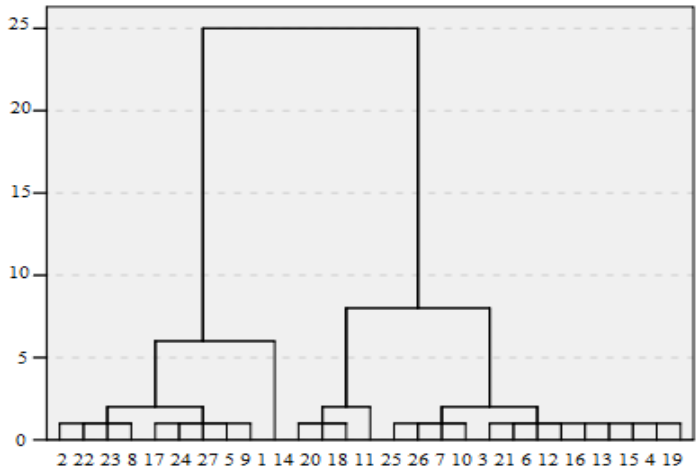

Fig. 1: Clustering analysis tree of red wine

divided the red wine samples into two categories. The first category contained samples 2, 5, 8, 9, 14, 17, 22, 23, 24 and 27 and the second category contained samples 1, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, 18, 19, 20, 21, 25 and 26, respectively. Values of all the physicochemical indicators but the color ones of the first category were greater than those of the second category. It showed that the first class of the red wine was relatively dark and the tone was darker. It had sour taste and rich aroma. However, the second class was lighter in color and partially brick red. It tasted thin and had less odor than the first class.

Result of TOPSIS comprehensive evaluation: We calculated the entropy weights of four principal components of the red wine. As shown in Table 5, iterative weights calculated by BP neural network had a small difference from weights before iterating, proving that weights determined by information entropy had a high accuracy.

We conducted TOPSIS comprehensive evaluation regarding the two categories of red wine samples. Below are their positive and negative ideal solutions.
First category of red wine:

Negative ideal solution: S^- = [0.0306 -0.0261 0.1156 0.0054]

Positive ideal solution: S^+ = [0.1743 0.0613 0.2038 0.0557]

Second category of red wine:

Negative ideal solution: S^- = [-0.0071 0.0289 0.0186 0.0162]

Positive ideal solution: S^+ = [0.1792 0.0792 0.0574 0.0706]

The values of the vector S^+ represented ideal solutions for various physicochemical indicators while S^- represented non ideal solutions. Each value in the vector was permuted in accordance with the order of main components. The first value of the vector represented the ideal solution of component 1 and the last represented the ideal solution of component 4.

The optimal value stood for the relative closeness to the ideal solution can be calculated from negative and positive ideal solution. To be exact, the optimal value stood for the quality of wine. The higher the score was, the better the wine was. The wine grading standard can be determined according to the optimal values and their distribution. According to the distribution of optimal values of wine samples, we divided values into three intervals. Meantime, every category of wine was divided into three levels. Each interval corresponded to a particular grade of wine quality. Grade I indicated the worst quality, while Grade III stood for the best quality. The grading standards of red wine are shown in Table 6.

Table 7 shows scores of wine samples. If optimal value was less than 0.5, then the solution of corresponding wine samples approached the negative solution. On contrast, if optimal value was more than 0.5, the solution of corresponding wine samples approached the positive solution. From Table 7, we found that most optimal values were less than 0.5. So we concluded that the whole qualities of most wine samples were generally not high. And the distance

Table 6: Grading standards of red wine

Grade	Optimal value of first category	Optimal value of second category
I	0.4 or less	0.3 or less
II	0.4~0.5	0.3~0.5
III	0.5 or more	0.5 or more

Table 7: Optimal values of wine samples

Grade	Optimal value of first category	Optimal value of second category
I	0.3522, 0.3670, 0.3713, 0.3736, 0.3748	0.1079, 0.2357, 0.2753
II	0.4113, 0.4644, 0.4294	0.4840, 0.3548, 0.3587, 0.4047, 0.4932, 0.4536, 0.4727, 0.4271, 0.4525
III	0.5700, 0.6441	0.7748, 0.5298, 0.5613, 0.5395, 0.7592

Table 8: Wine classification results

Grade	First category (%)	Second category (%)
I	50	17.65
II	30	52.94
III	20	29.41

between the value of the best wine and that of the worst wine in three categories were bigger than 0.5, which showed good discrimination of using TOPSIS method.

Based on optimal values, we graded the red wine according to the standard shown in Table 6. Table 8 shows wine classification results. In the first category of red wine, wine of grade I accounted for 50%, wine of grade II accounted for 30%, wine of grade III accounted for 20%; In the second category of red wine, wine of grade I accounted for 17.65%, grade II accounted for 52.94%, grade III accounted for 29.41%. On the whole, with red wine, the quality of the lighter category was slightly higher than the darker category.

CONCLUSION

We grasped the principal components of the physicochemical indicators using Principle component analysis. And then we calculated each main component weight based on the method of entropy weights. To verify the accuracy of the weights calculation, we used BP neural network model to iterate weights. The results of BP neural network showed that there were narrow difference between iterative weights and initial weights, which proved that weights determined by information entropy had a high accuracy. After weights accuracy was verified, we clustered red wine samples into two categories. Weighted cluster analysis worked well in clustering. We applied the weighted TOPSIS method to objectively evaluating the quality of various types of wine, which showed good discrimination in assessment of wine quality. The method has displayed good practicality and can be used in cases where there are no other objective criteria available. It steers clear of the thorny problem of determining subjective weights in general evaluation and conducts a comprehensive evaluation of the quality of the wine, playing an important role in the promotion of scientific, standardized and institutionalized evaluation of the wine quality. What's more, the model can be widely used in food and other quality evaluation.

ACKNOWLEDGMENT

This study is supported by the National Natural Science Foundation of China (No. 11301535).

REFERENCES

Aly, M., 2005. Survey on multiclass classification methods. Technical Report, Caltech, USA.

Bentlin, F.R.S., C.M.M.D. Santos, E.M.M. Flores and D. Pozebon, 2012. Lanthanides determination in red wine using ultrasound assisted extraction, flow injection, aerosol desolvation and ICP-MS. Anal. Chim. Acta, 710: 33-39.

Burin, V.M., L.L.F. Costa, J.P. Rosier and M.T. Bordignon-Luiz, 2011. Cabernet sauvignon wines from twom different clones, characterization and evolution during bottle ageing. LWT-Food Sci. Technol., 44: 1931-1938.

Chen, S.J. and C.L. Hwang, 1992. Fuzzy Multiple Attribute Decision Making: Methods and Applications. Springer-Verlag, Berlin.

Cruz-Ramiaírez, M., J.C. Fernández, J. Sánchez-Monedero, F. Fernández-Navarro, C. Hervá s-Martínez, P.A. Gutiérrez and M.T. Lamata, 2010. Ensemble determination using the TOPSIS decision support system in multi-objective evolutionary neural network classifiers. Proceeding of the 10th International Conference on Intelligent Systems Design and Applications. Cario, Egypt, pp: 513-518.

He, Y., X.L. Li and X.F. Deng, 2007. Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model. J. Food Eng., 79: 1238-1242.

Hernanz, D., A.F. Recámales, M.L**.** González-Miret, M.J. Gómez-Míguez, I.M. Vicario and F.J. Heredia, 2007. Phenolic composition of white wines with a prefermentative maceration at experimental and industrial scale. J. Food Eng., 80(1): 327-335.

Hwang, C.L. and K. Yoon, 1981. Multiple Attribute Decision Making: Methods and Applications. Springer-Verlag, Berlin.

Jiang, L., J. Xue, Y.J. Wang, Q. Lin and L.J. Du, 2008. Application of SNIF-NMR technique (site-specific natural isotope fractionation-nuclear magnetic resonance) in grape wine quality evaluation. Liquor-Making Sci. Technol., 169: 60-62.

Jin, W.Q., 2005. Fuzzy classification based on fuzzy association rule mining. Thesis, NC State University, Raleigh.

Kavuri, N.C. and M. Kundu, 2011. ART1 network: Application in wine classification. Int. J. Chem. Eng. Appl., 2(3): 189-195.

Kruzlicova, D., J. Mocak, B. Balla, J. Petka, M. Farkova and J. Havel, 2009. Classification of slovak white wines using artificial neural networks and discriminant techniques. Food Chem., 112(4): 1046-1052.

Lawless, H.T. and H. Heymann, 1999. Sensory Evaluation of Food: Principles and Practices. Springer-Verlag, Berlin.

Li, X.X., K.S. Wang, L.W. Liu, J. Xin, H.R. Yang and C.Y. Gao, 2011. Application of the entropy weights and TOPSIS method in safety evaluation of coal mines. Proc. Eng., 26: 2085-2091.

Osorio, D., J.R. Pérez-Correa, E. Agosin and M. Cabrera, 2008. Soft-sensor for on-line estimation of ethanol concentrations in wine stills. J. Food Eng., 87(4): 571-577.

Römisch, U., D. Vandev and K. Zur, 2006. Application of interactive regularized discriminant analysis to wine data. Aust. J. Stat., 35(1): 45-55.

Shoemaker, P.A., M.J. Carlin and R.L. Shimabukuro, 1991. Back propagation learning with trinary quantization of weights updates. Neural Networks, 4(2): 231-241.

Sun, Y.F., Z.S. Liang, C.J. Shan, H. Viernstein and F. Unger, 2011. Comprehensive evaluation of natural antioxidants and antioxidant potentials in *Ziziphus jujuba Mill. var. spinosa* (*Bunge*) *Hu ex H. F.* Chou fruits based on geographical origin by TOPSIS method. Food Chem., 124(4): 1612-1619.

Xie, Z.H., Y. Zhang and C. Jin, 2012. Prediction of coal spontaneous combustion in goaf based on the BP neural network. Proc. Eng., 43: 88-92.

Yoon, K.P. and C.L. Hwang, 1995. Multiple Attribute Decision Making. Sage Publication, Thousand Oaks, CA.

Zhang, Y.L., X.D. Liu and X.Y. Wang, 2009. A novel weighted fuzzy clustering analysis based on AFS theory. Proceeding of the 9th International Conference on Hybrid Intelligent Systems. Shenyang, China, 3: 346-350.