

# Application on sensory prediction of Chinese Moutai-flavour liquor based on ATR-FTIR

Wang Fan <sup>1</sup>, Shao Chunfu <sup>1,2,\*</sup>, Chen Qi <sup>1</sup>, Meng Tianyi <sup>3</sup>, and Li Changwen <sup>3</sup>

<sup>1</sup> Tasly Academy, Tasly Group, Tianjin 300410, China PR

<sup>2</sup> College of Biotechnology, Tianjin University of Science and Technology, Tianjin 300457, China PR

<sup>3</sup> Guotai Liquor Co. Ltd., Renhuai Guizhou 564501, China PR

**Abstract.** ATR-FTIR combined with chemometrics was applied to establish SVM classification models aiming to evaluate sensory quality of Chinese Moutai-flavour liquor. Transformation of ATR-FTIR data, selection of effective wavenumbers as well as determination of  $c$  and  $\gamma$  were performed in succession, while the verification of models was deployed applying unknown samples. Finally, taste-prediction models of raw grain and cleanliness have an accuracy reaching 90%. Model of after-taste has an accuracy of 80% and others are lower than 70%. As for some flavours, ATR-FTIR and chemometrics technology provided an effective method for quality analysis of Chinese Moutai-flavour liquor.

## 1 Introduction

Moutai flavour spirit, is a typical Chinese liquor with a soy sauce aroma, made from grains and *daqu*. During fermentation, the metabolism of the functional micro-organisms gives rise to highly stable heat- and acid-resistant enzymes such as amylases, proteases, glucoamylases, cellulases, glucosidases, xylanases as well as various dehydrogenases and phosphoenolpyruvate carboxykinases involved in oxidation-reduction reactions. All kinds of extreme micro-organisms (such as *Bacillus licheniformis* [1] and *Bacillus subtilis* [2-4], which produce the soy sauce aroma) and enzymes, encouraged by the long production process, help to produce rich flavour compounds (including esters, aldehydes, sulphur and nitrogen) of the liquor. These flavor compounds enter into the basic distillates by means of different processes and further assist in the formation of the unique taste and quality of Moutai flavour liquor which sensory characteristics are as follows: a) Smell: featuring soy Moutai-flavour aroma, fine and strong tasting, with enduring aroma lingering in empty glasses. b) Taste: tasting pleasantly strong and full, with lingering aftertaste. c) Style: featuring soy Moutai-flavour aroma, refined and delicate, pleasantly strong and full, with enduring aftertaste and lingering aroma in empty glasses [5].

The assessment of Chinese spirits is traditionally performed by professional tasters, which outcome mainly depends on the acute olfaction of tasters and this approach is time-consuming, labour intensive and affected by subjective factors or even biased. Therefore, researchers have been sought for sophisticated instruments combined with multiple analysis methods for decades. So far technologies such as GC [6], GC-MS

[7], GC-O [8] and FTIR[9], which are usually combined with chemometrics including PLS [10], SVM [11], ANN [12], CA, have been applied on quantitative or qualitative analysis of Chinese spirits to determine component, concentration [13], or to identify source, geographic origin [14], brand, authenticity [15] or storage time [16].

Qin Ouyang studied on prediction of the overall sensory score of Chinese rice wine applied NIR spectroscopy and Back propagation artificial neural network (BPANN) combined with adaptive boosting (AdaBoost) algorithm [17]. Daniel Cozzolino [18] investigated the relationship between sensory analysis and visible and near infrared spectroscopy in two Australian white wine varieties using PLS regression. The correlation coefficients were greater than 0.70 for estery, lemon and honey, and less than 0.50 for passionfruit, overall flavor and sweetness in both calibration and cross validation.

This paper aimed at predicting sensory quality of Moutai-flavour base liquor applying ATR-FTIR and SVM.

## 2 Experiment

### 2.1 Apparatus

Spectrometer (Tensor 27, Bruker Ltd. ), equipped with an attenuated total reflectance (ATR) accessory (PIKE, Ltd.).

### 2.2 Sample

\*Corresponding author: Shao Chunfu: [shaochunfu@tasly.com](mailto:shaochunfu@tasly.com)

Moutai-flavor base liquor (different batches from 2015-2018) with its sensory analysis provided by Guotai Liquor Co. Ltd.. Sensory quality was evaluated by means of a score from 0 to 10 for each flavour. Prediction models were established using data collected in 2015-2017 and data of 2018 were tested these models.

**Table 1** Number of samples used in model establishment and verification

Flavour	Model Establishment	Model Verification
Raw grain	169	50
grain	427	80
Sourness and astringency	595	120
Moutai-flavour	1611	300
Thickness	2207	400
Pungent	2207	400
After-taste	2207	400
cleanliness	2207	400

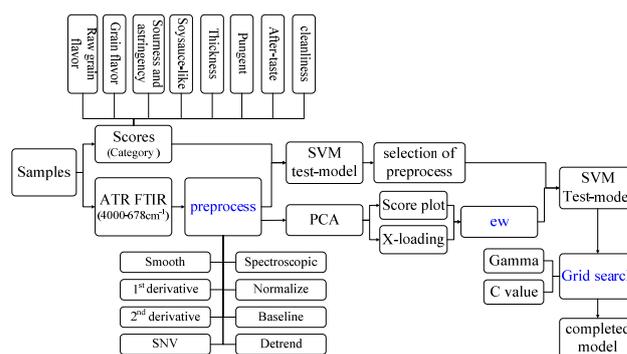
### 2.3. Methods

IR spectra were recorded with an accumulation of 32 scans in 4000-650  $\text{cm}^{-1}$  range with a resolution of 4  $\text{cm}^{-1}$ . The score of samples offered by panel of sensory analysis was set as one category to establish a Support Vector Machine (SVM) model. The establishment was performed on *The Unscrambler* (version 10.3, CAMO Software AS, Norway) and the infrared absorbance values were chosen as independent variables. Appropriate preprocess on infrared data should be selected according to the accuracy of test-models which are built respectively with different sets of parameters. Effective wave (ew) range is suggested by Principal Component Analysis (PCA) which x-loadings plot highlights regions of the most importance and performs as an indicator that helps to determine beneficial wavenumbers which contain the most significant information of the spectra. Finally, with the best set of parameters (preprocess and ew), a grid search is required for the best gamma and c value which are two significant parameters of Radial Basis Function (RBF) kernel SVM. C controls overfitting of model, and gamma controls the degree of nonlinearity of the model. Gamma is inversely related to sigma, which is a degree for spread around a mean in statistics: the higher the value of gamma, the lower the value of sigma, thus the less spread or the more nonlinear the behavior of the kernel. Finally, unknown samples would be predicted by the SVM model to confirm the accuracy of models.

**Table 2** Descriptions of preprocess on ATR FTIR

Transformation	Purpose and function
Normalize	“Scale” samples in order to get all data on approximately the same scale.
Standard normal variate transformation (SNV)	Remove slope variation to correct for scatter effects. Each spectrum $x_i$ is corrected individually by first centering the spectral values. Then the centred spectra are scaled by the standard deviation calculated from the individual spectral values.
Baseline offset	For each sample, the value of the lowest point in the spectrum is subtracted from all the variables. The result of this is that

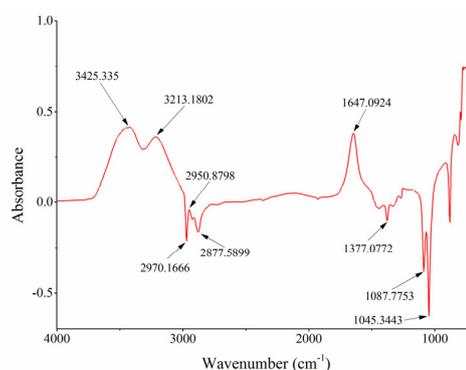
	the minimum value is set as 0 and the rest are positive values.
Detrend	Remove the effects of baseline shift and curvi-linearity.
Gap derivatives	Baseline drifts are reduced and slight spectral differences are enhanced by means of the computation of 1 <sup>st</sup> and 2 <sup>nd</sup> derivatives. To avoid enhancing the noise, which is a consequence of derivation, spectra are first smoothed.
Spectroscopic	Convert units from absorbance by taking the inverse logarithm of it to give reflectance or transmittance respectively.



**Fig. 1** Illumination of this study

## 3 Results and discussion

### 3.1 ATR-FTIR spectra



**Fig. 2** ATR-FTIR spectra of sample collected applying Bruker Tensor 27 spectrometer

Chinese spirit contains water, ethanol and flavouring components. But it is the flavouring components which have a low proportion of 1%~2% that reflect the aroma, taste and quality of liquor product. So ethanol was deducted as background when spectra collected, which caused the negative peaks (Shown in Fig. 2). FTIR spectrum reflects holistic characteristics of sample, which contains overall information of components that spectral data were regarded as multivariable and put into software to analyse the relationship between FTIR information and sensory quality which is represented by scores.

**Table 3** Negative peaks

Peak( $\text{cm}^{-1}$ )	Identification
1087,1047	stretching vibration of C-OH
2878	symmetric stretching of $\text{CH}_3$
1378	symmetric swing of $\text{CH}_3$

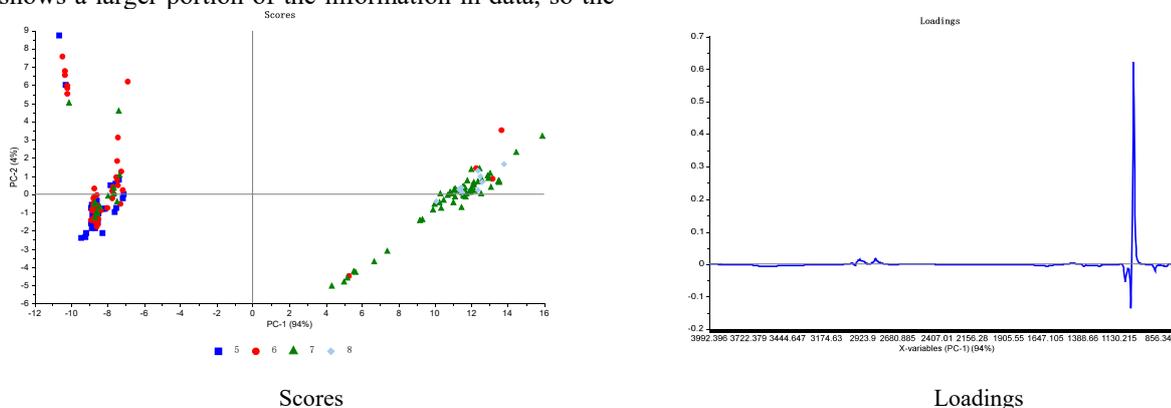
### 3.2 Optimization of model

Taste-prediction models were established respectively after a series of selections of preprocess, ew and SVM parameters.

As for raw grain, preprocess of *baseline* or *Spectroscopic* on FTIR could lead to better validation accuracy. Then a PCA based on the transformed FTIR was performed to select ew. Scores plot and loading plot were shown as Fig. 3. The closer the samples are in the scores plot, the more similar they are with respect to the two components concerned. The plot can be used to interpret differences and similarities among samples and help to determine which variables are responsible for differences between samples. The sum of the explained variances for the 2 components is 98% and the plot shows a larger portion of the information in data, so the

relationships can be interpreted with a high degree of certainty. PCA is a good way to detect important variables.

Samples were divided into two parts by the ordinate axis. The vast majority of samples which scores were 5 and 6 are located in second and third quadrants and those with 7 and 8 are most in first and fourth quadrants. This distribution indicates that PC1 is a major factor distinguishing higher scores and lower scores but an apparent trend between 5 and 6 or 7 and 8 wasn't be found. Then loadings plot on PC1 was checked to find ew and 1122-920cm<sup>-1</sup> was a significant range than others for classification owing to its high values on Y-axis. Then a grid search, which was an orthogonal test from 0.01 to 100 on 5 levels, was conducted to seek optimal gamma and c value. When c was 100 and gamma was 0.01, model has highest validation accuracy, so this combination was selected for SVM model of raw grain flavour.



**Fig. 3** PCA on FTIR data with *Spectroscopic* transformation

**Table 4** Result of models produced under different parameters

preprocess	SVM		EW (cm <sup>-1</sup> )	SVM ew	
	Training accuracy (%)	Validation accuracy (%)		Training accuracy(%)	Validation accuracy(%)
RAW	63	63			
1 <sup>st</sup> derivative	43	43			
1 <sup>st</sup> derivative + SNV	71	60	1200-678	92	55
2 <sup>nd</sup> derivative	43	43			
2 <sup>nd</sup> derivative + SNV	76	57	1137-921	69	65
Normalize(Area)	43	43			
Normalize(Mean)	71	62	1200-800	66	65
Baseline	63	63	1200-900	63	63
SNV	63	64			
SNV+detrend	63	63			
Spectroscopic	63	63	1122-920	69	67
Detrend	63	63			
Detrend+SNV	63	63			

**Table 5** Validation accuracy of models under different parameters

Gamma	C					
	0.01	0.1	1	10	100	
0.01	43.1953	63.9053	63.9053	64.4970	68.6391	
0.1	43.1953	63.3136	64.4970	63.3136	61.5385	
1	43.1953	55.6213	63.5385	56.8047	56.2130	
10	43.1953	43.1953	53.8462	51.4793	49.1124	
100	43.1953	43.1953	44.9704	44.3787	43.1953	

**Table 6** All taste models for sensory prediction established by ATR-FTIR and SVM

Flavour	Transformation	Range(cm <sup>-1</sup> )	Gamma	C	Training accuracy	Validation accuracy
Raw grain	Spectroscopic	1122-920	0.01	100	72.19%	68.05%
Grain	1 <sup>st</sup> derivative + SNV	1200-678	0.1	1	89.70%	89.23%
Sourness & astringency	2 <sup>nd</sup> derivative + SNV	1800-678	0.01	1	72.15%	63.43%
Cleanliness	RAW	4000-678	100	100	100%	95.05%
Moutai-flavour	1 <sup>st</sup> derivative + SNV	1200-678	0.01	1	72.42%	66.66%
Thickness	SNV + detrend	3700-2889, 1800-678	0.1	1	80.51%	77.00%
Pungent	Spectroscopic	1100-900	1	10	80.13%	69.41%
After-taste	Spectroscopic	1145-890	0.1	100	85.06%	81.79%

Other flavour models were established according to the above process, which performance was shown in

Table 6. Models of *Cleanliness*, *Grain*, *After-taste* have better predictive capability which accuracy reach 80%. Some flavours cannot be evaluated precisely and their models need to be updated and corrected with the large amount of experimental data continue to accumulate.

### 3.3 External validation

Additional samples were predicted as unknown samples by the completed models to confirm accuracy (the consistency between predicted and sensory scorings). The results were listed in Table 7. It is indicated that *grain* and *cleanliness* have better accuracy reaching 90%. *After-taste* has an accuracy of 80% and others are lower than 70%.

**Table 7** Result of external validation

flavour	accuracy
Raw grain	70%
Grain	90%
Sourness & astringency	70%
Cleanliness	96%
Moutai-flavour	66%
Thickness	63%
Pungent	69%
After-taste	80%

### Acknowledgement

This work was supported by Guizhou Provincial Commission of Economy and Information Technology 2016016.

### References

1. R. Zhang, Y. Xu. *Industrial Microbiology*. **40**, 1-7 (2010)
2. B. Zhu, Y. Xu, W. Fan. *J Biosci Bioeng*. **108**, S122 (2009)
3. B. Zhu, Y. Xu, W. Fan. *New Biotechnol*. **25S**, 237 (2009)
4. B. Zhu, Y. Xu. *J Ind Microbiol Biot*. **37**, 179-186 (2010)

5. Y. Xu, K. Ji. *Alcoholic Beverages*, Woodhead Publishing, (<http://www.woodheadpublishing.com>, 2012).
6. Z. Xiao, J. Luo, Y. Niu, M Wu. *Nat Prod Res*. **32**, 1-6 (2017)
7. Z. Xiao, D. Yu, Y. Niu, F Chen, S Song. *J Chromatogr B*. **945-946**, 92-100 (2014)
8. Z. Xiao, J. Chen, Y. Niu, F Chen. *J Chromatogr B*. **1063**, 226 (2017)
9. C. Li, J. Wei, Q. Zhou, S Sun. *J Mol Struct*. **883-884**, 99-102 (2008)
10. P. S. Sampaio, A. Soares, A. Castanho, A.S. Almeida, J Oliveira, C Brites. *Food Chem*. **242**, 196-204 (2018)
11. H. Chen, C. Tan, T. Wu, L Wang, W Zhu. *Spectrochim Acta A*. **130**, 245-249 (2014)
12. G. Astray, B. Gullón, J. Labidi, P Gullón. *Ind Crop Prod*. **92**, 290-299 (2016)
13. W. Zhu, C. Ren, Y. Nie, Y Xu. *Food Control*. **64**, 37-44 (2016)
14. H. Qin, D. Huo, L. Zhang, L Yang, S Zhang, M Yang, C Shen, C Hou. *Food Res Int*. **45**, 45-51 (2012)
15. J. Han, Z. Li, P. Shi, F Li, Y Duan. *Sens Actuators B Chem*. **248**, 101-107 (2017)
16. Q. Peng, X. Xu, W Xing, B Hu, C Shen, R Tian, X Li, Q Xu, J Chen, F Chen, H Zou, G Xie. *Innov Food Sci Emerg*. **44**, 212-216 (2017)
17. Q. Ouyang, Q. Chen, J. Zhao. *Spectrochim Acta A*. **154**, 42-46 (2016)
18. D. Cozzolino, H. E. Smyth, K. A. Lattey, W Cynkar, L Janik, R. Damberg, I.L. Francis, M Gishen. *Anal Chim Acta*. **539**, 341-348 (2005)