

Application of Radial Bases Function Network and Response Surface Method to Quantify Compositions of Raw Goat Milk with Visible/Near Infrared Spectroscopy

Ching-Lu Hsieh¹, Chao-Yung Hung², Mei-Jen Lin³

¹ Associate professor, Department of Biomechatronics Engineering, National Pingtung University of Science and Technology, Taiwan, R.O.C. Corresponding author.

chinglu@mail.npust.edu.tw, Tel: 886-8-7703202#7583

² Graduate student, Department of Biomechatronics Engineering, National Pingtung University of Science and Technology, Taiwan, R.O.C.

³ Associate professor, Department of Animal Science, National Pingtung University of Science and Technology. Taiwan, R.O.C.

Abstract. Raw goat milk pricing is based on the milk quality especially on fat, solid not fat (SNF) and density. Therefore, there is a need of approach for composition quantization. This study applied radial basis function network (RBFN) to calibrate fat, SNF, and density with visible and near infrared spectra (400~2500 nm). To find the optimal parameters of goal error and spread used in RBFN, a response surface method (RSM) was employed. Results showed that with the optimal parameters suggested by RSM analysis, R^2 difference for training and testing data set was the smallest which indicated the model was less possible of overtraining or undertraining. The R^2 for testing set was 0.9569, 0.8420 and 0.8743 for fat, SNF and density, respectively, when optimal parameters were used in RBFN.

Keywords: Raw goat milk, Radial basis function network, Response surface method, NIR

1 Introduction

Goat milk provides abundant nourishment that has been treated as an important daily supply especially for the people at the stage of recovery or youth. Price of goat milk usually higher than cow milk either in raw or fresh format. Dairy factory evaluates milk price based on its quality, such as the ratio of fat, SNF (Solid Not Fat), and density, which drives dairy farmer's attention on monitoring the constituents of milk. Therefore, a quick method for measuring goat milk composition is highly needed.

For decades, near infrared (NIR, 700~25000 nm, or 0.7~25 μm) or infrared (IR, 25~50 μm) has been proven high performance in qualitative and quantitative analysis for agriculture, food, and pharmacy industry. For milk quality evaluation and quantization, AOAC [1] suggested that some milk components are good correlated to IR energy absorption, such as, CH groups in fatty acid chains of fat molecules absorb at 3.48 μm . Many studies have been conducted on cow milk [2,3,4], but studies on

raw goat milk are comparatively less, meanwhile most approaches used in previous researches usually needs chemometrics that users usually hard to follow the theory.

Artificial neural networks (ANN) have been extensively used in many fields because it can perform a curve-fitting operation in a multidimensional space even it has been thought as a black box. After properly training, ANN model can be applied straightforwardly. Thus, it is a good option for dairy farmers for online application. One drawback of ANN is that they are highly nonlinear in the parameters and their parameter estimation may be trapped at a local minimum in learning process. One alternative to highly nonlinear-in-the-parameter neural networks is the radial bases function network (RBFN) [5]. The performance of an RBFN depends upon the choices of the number and the centers of hidden units. Most choice of centers is to let each data point in the training set correspond to a center. In this case the number of degree of freedoms in the network is equal to the number of training data. In some cases, such as data contaminated by noise, then the overfitting phenomenon will occur. One of possible way to access overfitting condition is to separate samples into training data set and testing data set, then compare the performance on each set. In this study, we used response surface method (RSM) to tune network parameters so as to avoid or lessen overtraining or undertraining problem. Therefore, this article has three purposes:

1. obtain visible/NIR spectroscopy of raw goat milk samples mixed with 0%, 2%, 5%, 10%, 20%, 30%, 50%, 70% and 100% reconstituted milk to vary sample compositions
2. apply RBFN to quantify the compositions of the goat milk, i.e., fat, SNF, and density
3. use RSM to study the effect of RBFN parameters, i.e., goal error and spread so as to find the optimum model parameters

2 Materials and Methods

2.1 Sample preparation

Raw goat milk samples were collected in the morning from a dairy farm near our campus, and they were sent to our laboratory within 2 hr. Sampling process was conducted from Oct. 2010 to Nov. 2010. In order to have different composition, reconstituted goat milk were made by diluting powdered goat milk with water on the basis of protein of raw milk. It means the reconstitutes goat milk samples have as close as protein ratios to raw goat milk samples. After the reconstituted milk was prepared, it was mixed with raw milk in volume ratio of 0:100 (0%), 2:98 (2%), 5:95 (5%), 10:90 (10%), 20:80 (20%), 30:70 (30%), 50:50 (50%), 75:25 (75%), and 100:0 (100%). Thus total 324 test samples were prepared.

Each sample was scanned to record its visible/NIR spectrum. Sample was also analyzed by milk composition analyzer (Expert, Scopo Electric, Europe) on fat, protein, SNF (solid-not-fat), lactose, and density. But, only fat, SNF, and density were studied in here. Monochrometer (FOSS NIRS 6500, NIRSYSTEM, US) was used to

obtain spectrum (400~2498 nm at 2 nm intervals). From all tested samples, 216 samples were randomly selected as training set while the remaining samples (108) were used as testing set.

To correct the baseline shift and to smooth noise on spectral data, each spectrum was proceed with baseline treatment, Savitzky–Golay smoothing with 4th order 11 data points, and standard normal variate scaling. These data processes were supported by a PLS-Toolbox (Eigenvector co.) compatible to MATLAB.

2.2 Radial Basis Function Networks

A RBFN is one whose output is symmetric around an associated center, c_i . That is, $\phi(x) = \phi(\|x - c_i\|)$, where $\|\cdot\|$ is a vector norm or distance, such as, Euclidean distance. A radial basis neuron with input p ($R \times 1$ dimension) and with a bias b , then its output can be formulated as $f(\|w - p\|b) = f(n)$. One transfer function f of radial basis neuron support by MATLAB is

$$\text{radbas}(n) = e^{-n^2} \quad (2)$$

A RBFN is a supervised network and it takes matrices of input vectors P and target vectors T , and a spread constant for the radial basis layer. The spread determines the width of an area in the input space to which each neuron responds. Thus spread should be large enough to overlap the regions of input space. Thus, designing an RBN involves selecting the goal error and width of RBFN. In here, RBFN function was supported by Neural Network Toolbox (MATLAB).

2.3 Response surface method

Response y are affected by variables $\xi_1, \xi_2, \dots, \xi_k$ that are usually called the natural variables. In many RSM works it is convenient to transform the natural variables to coded variables x_1, x_2, \dots, x_k . In terms of the coded variables, the response function can be written as [6]

$$\eta = f(x_1, x_2, \dots, x_k) \quad (3)$$

For the case of two independent variables, the second order model is

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 \quad (4)$$

Many applications of RSM involve fitting and checking the adequacy of a second order model. An experimental design called central composite design (CCD) has been widely used for fitting a second-order response surface. Natural variables and coded variables of each CCD observation is listed in Table 1 suggested by previous run. Model performance of RBN was evaluated by coefficient of determination R^2 for training set (R_c^2) and test set (R_v^2). These analyses were conducted by MATLAB. Thus, RBFN was trained with each error goal and spread parameter listed on Table 1.

Table 1. Central composite design for RBN

<i>Observation</i>	<i>Error Goal</i>	<i>Spread</i>	<i>Coded variable</i>	
	ξ_1	ξ_2	x_1	x_2
1	0.001	0.5	-1	-1
2	0.1	0.5	1	-1
3	0.001	1.5	-1	1
4	0.1	1.5	1	1
5	0.000373	1	-1.414	0
6	0.3726	1	1.414	0
7	0.01	0.3344	0	-1.414
8	0.01	1.707	0	1.414
9	0.01	1	0	0
10	0.01	1	0	0
11	0.01	1	0	0
12	0.01	1	0	0

3 Results and Discussion

3.1 RSM and RBN performance

Based on the CCD experiment, 12 observations in Table 1 were examined by RBFN (MATLAB) and each corresponding R^2 for training set (R_c^2) and test set (R_v^2) are shown in Table 2. The highest R_c^2 for fat, SNF, and density are 0.9943, 0.9917, and 0.9902, respectively, while the highest R_v^2 are 0.9549, 0.8208, and 0.8599. R_v^2 is smaller than the corresponding R_c^2 , therefore overtraining problem might exist.

Table 2. The performance of RBFN at each central composite design observation

Obs.	Coded variable		Fat		SNF		Density	
	X_1	X_2	R_c^2	R_v^2	R_c^2	R_v^2	R_c^2	R_v^2
1	-1	-1	0.9847	0.9492	0.9778	0.8209	0.9754	0.8599
2	1	-1	0.3376	0.4460	0.0994	0.0546	0.2095	0.1603
3	-1	1	0.9850	0.9506	0.9779	0.7755	0.9739	0.7751
4	1	1	0.3698	0.4598	0.1146	0.1009	0.2332	0.1882
5	-1.414	0	0.9943	0.9549	0.9917	0.6872	0.9902	0.7392
6	1.414	0	0.2295	0.2267	0.0133	0.0010	0.0746	0.0122
7	0	-1.414	0.8538	0.8587	0.7817	0.6321	0.7639	0.6341
8	0	1.414	0.8521	0.8617	0.7841	0.7045	0.7379	0.6171
9	0	0	0.8585	0.8955	0.7855	0.7367	0.7526	0.7372
10	0	0	0.8585	0.8955	0.7855	0.7367	0.7526	0.7372
11	0	0	0.8585	0.8955	0.7855	0.7367	0.7526	0.7372
12	0	0	0.8585	0.8955	0.7855	0.7367	0.7526	0.7372

R_c^2 is for training data set, and R_v^2 is for test data set.

Three conditions of goal error and spread suggested in the optimal regions were set on the RBFN and run the model again. Results show the highest R_v^2 is 0.9569, 0.8420 and 0.8743 for fat, SNF and density, respectively. The corresponding R_c^2 is 0.9573, 0.9181, and 0.9255 respectively. These parameters show less problem on model overtraining or model undertraining because they have the lowest difference between R_c^2 and R_v^2 . The epochs for these optimal tests were 25 for fat, 50 for SNF, and 25 for density. The numbers of centers of these optimal RBFN were 37 for fat, 63 for SNF, and 49 for density. Scatter distribution of 108 testing samples on fat for observation and prediction are shown in Fig. 1. Model used is the RSM optimal model. Results shown the prediction error randomly occurred that made the data points distribute randomly beneath or above the identical line.

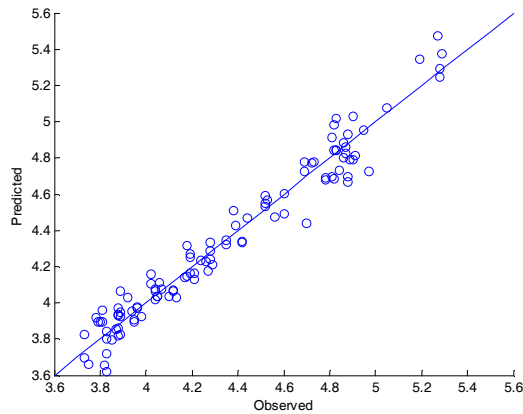


Fig. 1. Scatter plot of observation and prediction of testing set for fat (%) at the optimal parameters of goal error 0.0028 and spread 1.0

4 Conclusion

Parameters of goal error and spread of radial basis function networks (RBFN) were tested under the assistance of response surface method (RSM). The optimal parameters were found and validated for three raw goat milk compositions: fat, solid not fat (SNF), and density. Result shows that RSM suggested the optimal parameters for fat in RBF were goal error 0.0028 and spread 1.0, while for SNF were goal error 0.0037 and spread 0.9 and for density were goal error 0.0028 and spread 0.7. With these optimal parameters the R_v^2 was 0.9569, 0.8420 and 0.8743 for fat, SNF and density, respectively.

Acknowledgments. This study was finally supported by Council of Agriculture, the Executive Yuan, Taiwan, R. O. C.

References

1. AOAC official methods of analysis. 1990., 972.16 Fat, lactose, protein, and solids in milk. Mid-infrared spectroscopic method. Association of Official Analytical Chemists, Arlington, VA. 816-818
2. Schmilovitch, Z., Maltz, E. and Austerweill, M. 1992. Fresh raw milk composition analysis by NIR spectroscopy. Proceedings of the International Symposium on Prospects for Automatic Milking. Wageningen. Netherlands. EAAP Publication No.65, 1992. Eds: A.H. Ipema, A.C Lippus J.H.M Metz & W. Rossing.
3. Martin, I. G., J. M. Hierroa, R. M. Sanchoa, J. S. Estebana, A. V. Quintanab, I. Revilla. 2007. Determination of the percentage of milk (cow's, ewe's and goat's) in cheeses with different ripening times using near infrared spectroscopy technology and a remote reflectance fibre-optic probe. *Analytica Chimica Acta* 604(2007):191-196
4. Mouazen, A.M., S. Dridi, H. Rouissi, J. De Baerdemaeker and H. Ramond. 2007. Feasibility study on using visible-near infrared spectroscopy coupled with factorial discriminant analysis technique to identify sheep milk from different genotypes and feeding systems. *Journal of Near Infrared Spectroscopy* 15(6):359-369
5. Chen, S., C.F.N. Cowan, and P.M. Grant. 1991. Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks. *IEEE Transactions on Neural Networks* 2(2): 302-309.
6. Myers, R. H. and D. C. Montgomery. 1995. *Response Surface Methodology*. John Wiley & Sons. P.1-78.