Journal of Computational Science & Engineering 2 (2011) 69-78

Journal of Computational Science & Engineering

Available at http://www.asocse.org ISSN 1710-4068

QSPR models for the prediction of odor threshold of aliphatic alcohols

Xuan Xu, Feng Luan *, Huitao Liu, Yuan Gao

College of Chemical and Chemical Engineering, Yantai University, Yantai 264005, P. R. China

Article Information

Article history:
Received 12 March 2011
Revised 10 May 2011
Accepted 25 May 2011
Available online 10 June 2011

Keywords:
Odor threshold
Aliphatic alcohol
MLR
RBFNN
QSPR

Abstract

Quantitative structure-property relationship (QSPR) models, which could predict odor threshold of aliphatic alcohols, were developed by applying chemical descriptors computed with quantum chemical PM3 algorithm and using linear and nonlinear method, respectively. Both of the models gave acceptable results. The model of nonlinear radial basis function neural networks (RBFNN) leads to a squared correlation coefficient (R^2) of 0.76 and root-mean-square error (RMS) of 0.528 for the test set and the values for linear model are 0.62 and 0.689 respectively. The QSPR models provide a rapid, simple and valid way to predict the odor threshold of aliphatic alcohol.

1. Introduction

The presence of airborne chemicals known as volatile organic compounds (VOCs) in the environment, including the home and the workplace, can lead to sensory irritation of the upper airways, a widely cited health effect from polluted indoor air [1, 2]. As is well known, VOCs also play a major role in the formation of various secondary pollutants through photochemical reactions in the presence of sunlight and nitrogen oxides [3, 4]. The number of industrial chemicals exceeds 100,000, of which perhaps one-third (30,000) could be classified as volatile organic compounds. It is clearly not economical to obtain sensory irritation data either directly on humans or indirectly from an animal assay for more than a very small proportion of VOCs that could be encountered in everyday life. A key issue in understanding human chemosensory perception involves knowledge of the relevant structural and physicochemical properties of chemicals that govern potency, i.e., absolute detection, and perceptual quality.

To better understand the sensory irritation of chemicals, odor threshold value is important. The odour threshold value of a volatile compound is defined as the minimum concentration at which the compound can be detected by the sense of

*Corresponding author Tel.: +86 535 6902063, Fax: +86 535 6902063
E-mail address: fluan@sina.com

smell. The odour threshold value of aliphatic alcohols is important since alcohols are widely used in food, perfume, cosmetic, detergents and other industries [5, 6].

There have been a number of correlations of odor detection thresholds (ODT) with various properties of odorant. The study by Laffort and Patte was the first to employ a physicochemical analysis [7]. Mihara and Masuda used a two-term regression equation to model the logarithm of the odor threshold of 60 disubstituted pyrazines [8]. Seeman et al. studied the odor profile of structurally similar pairs of 1,3-dialkylbenzenes and 2,6-dialkylpyridines as a function of the accessibility of the nitrogen atom and steric hindrance [9]. Winter related the activity of a series of ambergris-type odorants to a minimum accessible surface area about the ether oxygen in the molecule [10]. Edwards and Jurs also used discriminant analysis to study the ability of odorant molecules to stimulate activity of the enzyme adenylate cyclase [11]. Latter, computer assisted statistical methods have been used to study the odor thresholds of two sets of odor active molecules by the same authors. One data set included 53 aliphatic alcohols; the second data set included 74 mono and di-substituted pyrazine derivatives [12]. Chastrette has reviewed work up to 1996 [13]. Yamanaka showed that odour thresholds for several homologous series could be correlated with the odorant activity coefficient in water [14]. Abraham performed a model for odour thresholds for a series of 64 compounds, including esters, aldehydes, ketones, alcohols, carboxylic acids, aromatic hydrocarbons, terpenes and some of other VOCs [15].

All of previous studies attempt to produce an easy, accurate, and predictive model using different descriptors or methods. However, some models have been developed for relatively small data set of compounds. Most study attempts to produce the model with physic-chemical properties，not using the calculated molecular descriptors. Additionally, previous studies scarcely use none linear statistical technique to build model. The aim of the present work is to devise quantitative structure-property relationships (QSPRs) that could be used to correlate odor thresholds with relevant molecular descriptors calculated from chemical software alone by both linear and none linear methods, and thereby to approach prediction of such thresholds. The structural factors affecting the compounds' odor thresholds values were also investigated.

2. Materials and methods

2.1. Dataset

The odor threshold data for the 97 aliphatic alcohols was collected from the work of Schnabel *et al.*[16].

Concentration units of the experimental odor t hreshold were ppm. The data of the mean values of the reported range were transformed in logarithmic unit to linearize the experimental range of variation. The data set was split into a training set and a prediction set. The prediction set of 19 alcohols was selected randomly from the original 97 compounds with the remaining compounds constituting the training set. The training set of 78 compounds was used to adjust the parameters of the model, and the test set of 19 compounds was used to evaluate its predictive ability. A complete list of the compounds' name, their data set and corresponding LogT was shown in Table 1.

2.2. Generation of the descriptors

The calculation process of the molecular descriptors was described as below: molecules were drawn with Hyperchem and then pre-optimised using MM^{+} molecular mechanics force field [17]. A more precise optimisation was then done with the semi-empirical PM3 method in MOPAC6.0 [18]. All calculations were carried out at a restricted Hartree-Fock level with no configuration interaction. The molecular structur-es were optimised using the Polak-Ribiere algorithm until the root-mean -square gradient reached 0.001. The resulting geometry was

transferred into software CODESSA that can calculate constitutional, topological, electrostatic, and quantum chemical descriptors [19, 20]

Table 1 Chemicals, experimental, MLR and RBFNN predicted LogT

No.	Chemical Name	Log T		
		Experimental	MLR	RBFNN
	Training set			
1	Methanol	1.52	1.323	1.733
2	Propan-1-ol	3.96	3.553	3.677
3	Propan-2-ol	3.04	2.929	3.342
4	Butan-1-ol	4.77	4.535	4.308
5	Butan-2-ol	3.47	3.339	3.390
6	2-Methylpropan-2-ol	3.87	3.611	3.841
7	Pentan-1-ol	4.54	5.023	5.145
8	Pentan-2-ol	3.89	5.120	5.130
9	Pentan-3-ol	4.11	4.251	5.007
10	3-Methylbutan-1-ol	4.69	4.543	4.916
11	2-Methylbutan-2-ol	3.54	4.486	5.031
12	3-Methylbutan-2-ol	5.19	4.681	5.096
13	2,2-Dimethylpropan-1-ol	4.79	5.196	4.721
14	Hexan-2-ol	5.70	5.569	5.833
15	Hexan-3-ol	4.86	5.039	5.289
16	2-Methylpentan-1-ol	5.01	5.112	5.057
17	3-Methylpentan-1-ol	4.75	4.873	5.003
18	2-Ethylbutan-1-ol	5.75	5.261	5.305
19	2,2-Dimethylbutan-1-ol	4.24	4.770	4.275
20	2,3-Dimethylbutan-1-ol	4.86	5.020	4.740
21	3-Methylpentan-2-ol	4.74	5.145	4.392
22	4-Methylpentan-2-ol	5.59	4.972	4.542
23	2,3-Dimethylbutan-2-ol	5.59	5.220	4.657
24	3,3-Dimethylbutan-2-ol	4.60	5.304	4.666
25	3-Methylpentan-3-ol	4.51	4.738	4.996
26	Heptan-1-ol	5.43	5.659	5.499
27	Heptan-2-ol	6.31	5.810	5.488
28	Heptan-3-ol	5.57	5.481	5.486
29	5-Methylhexan-1-ol	5.38	5.601	5.498
30	2,2-Dimethylpentan-1-ol	6.81	6.400	7.287
31	2,3-Dimethylpentan-1-ol	5.00	5.480	5.492
32	2,4-Dimethylpentan-1-ol	5.58	5.461	5.225
33	3-Methylhexan-2-ol	5.81	5.642	5.526
34	5-Methylhexan-2-ol	5.40	5.588	5.455

35	3-Ethylpentan-2-ol	5.55	5.744	5.357
36	2,3-Dimethylpentan-2-ol	4.65	5.300	5.489
37	4,4-Dimethylpentan-2-ol	6.31	6.236	6.306
38	2-Methylhexan-3-ol	6.28	5.779	5.543
39	3-Methylhexan-3-ol	6.28	5.623	5.533
40	4-Methylhexan-3-ol	5.63	5.594	5.532
41	3-Ethylpentan-3-ol	6.29	5.471	5.578
42	2,2-Dimethylpentan-3-ol	6.31	6.135	6.477
43	2,3-Dimethylpentan-3-ol	5.29	5.178	5.405
44	2,4-Dimethylpentan-3-ol	5.80	5.492	5.469
45	Octan-2-ol	6.86	5.940	5.773
46	Octan-3-ol	5.97	5.754	5.856
47	Octan-4-ol	5.36	6.106	5.827
48	2-Methylheptan-2-ol	5.72	6.125	5.881
49	2-Methylheptan-3-ol	5.48	6.164	5.845
50	3-Methylheptan-3-ol	6.86	5.915	5.966
51	4-Methylheptan-3-ol	5.63	5.955	5.925
52	5-Methylheptan-3-ol	6.37	5.554	5.724
53	4-Ethylhexan-3-ol	6.06	5.906	5.911
54	2-Ethylhexan-1-ol	5.07	6.198	5.820
55	2,3-Dimethylhexan-2-ol	5.39	5.845	5.952
56	2,5-Dimethylhexan-2-ol	4.96	5.752	5.662
57	2,2-Dimethylhexan-3-ol	7.76	6.714	7.491
58	2,3-Dimethylhexan-3-ol	6.86	5.992	5.997
59	2,4-Dimethylhexan-3-ol	5.61	5.764	5.824
60	2,5-Dimethylhexan-3-ol	5.20	5.719	5.666
61	2,2,4-Trimethylpentan-3-ol	6.86	6.613	6.758
62	2,2,4-Trimethylpentan-1-ol	5.35	5.311	5.645
63	2-Methylheptan-4-ol	5.36	6.085	5.792
64	3-Methylheptan-4-ol	5.86	5.814	5.871
65	Nonan-1-ol	6.39	6.099	6.064
66	Nonan-2-ol	6.32	6.139	6.049
67	Nonan-3-ol	6.00	5.877	6.102
68	Nonan-4-ol	5.88	6.440	6.055
69	Decan-1-ol	7.33	6.334	6.261
70	Decan-2-ol	6.70	6.277	6.262
71	Decan-3-ol	5.95	6.087	6.357
72	Undecan-1-ol	5.96	6.484	6.417
73	Undecan-3-ol	6.96	6.374	6.336
74	Undecan-4-ol	6.47	6.357	6.410
75	Undecan-5-ol	5.47	6.569	6.440
76	Undecan-6-ol	6.96	6.673	6.423
77	Dodecan-2-ol	6.51	6.538	6.581

78	Dodecan-3-ol	6.49	6.479	6.626
	Test set			
1	2-Methylpropan-1-ol	3.77	4.330	4.347
2	2-Methylbutan-1-ol	5.44	4.370	5.039
3	Hexan-1-ol	4.90	5.419	5.008
4	4-Methylpentan-1-ol	4.75	5.039	5.219
5	2-Methylpentan-2-ol	4.25	5.517	4.319
6	2-Methylpentan-3-ol	5.24	5.059	5.077
7	Heptan-4-ol	5.30	5.675	5.502
8	2-Methylhexan-2-ol	4.70	5.679	5.570
9	2,4-Dimethylpentan-2-ol	4.57	5.903	5.546
10	5-Methylhexan-3-ol	5.31	5.065	5.146
11	Octan-1-ol	6.34	5.923	5.811
12	6-Methylheptan-2-ol	6.02	6.039	5.855
13	6-Methylheptan-3-ol	5.61	5.726	5.835
14	3,4-Dimethylhexan-2-ol	4.84	5.765	5.847
15	3,5-Dimethylhexan-3-ol	5.42	5.610	5.875
16	4-Methylheptan-4-ol	5.28	6.158	5.878
17	Nonan-5-ol	6.00	6.336	6.070
18	Undecan-2-ol	6.96	6.578	6.200
19	Dodecan-1-ol	5.98	6.704	6.476

In the present work, 388 descriptors were provided. Of them, 38 are constitutional, 38 are topological, 12 are geometrical, 73 are electrostatic and 220 are quantum chemical descriptors. In addition, the other 7 descriptors (approximate-surface-area, LogP, refractivity, polarizability, hydration energy, volume and mass) were calculated by Hyperchem and then added to the descriptors pool.

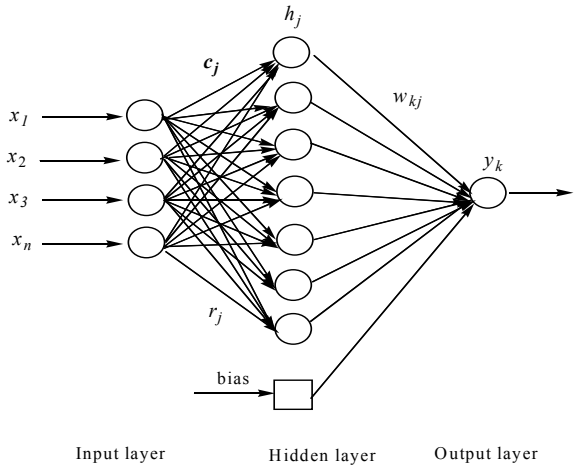

Figure 1. The architecture of RBFNN

3. Methodology

3.1 Theory of multiple linear regressions (MLR)

In quantitative structure property or activity relationships, molecular descriptors are correlated with one or more response variable. If it is assumed that the relationship is well represented by a model that is linear in the regressed variables, a suitable model may be:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \mathrm{K} + b_n X_n$$

In this equation, Y is the property, that is, the dependent variable; X_1 to X_n represent the specific descriptors, while b_1 to b_n represent the coefficient of those descriptors, and b_0 is the intercept of this equation.

A single MLR model was developed using the CODESSA software. The MLR model was built using training set and validated using an external prediction set. As is well know, MLR can't be used to model complex data, since in most cases the number of explanatory variables exceeds the number of objects. Therefore, it is often used in

combination with the stepwise procedure for variable selection [21, 22]. In this study the number of explanatory variables is more than the number of samples, and because of the fact that a

No.	Descriptor	Coefficient	Standard error	*t*-test
0	Intercept	227.59	67.68	3.363
1	Average Complementary Information content (order 0)	0.68	0.31	2.226
2	Approximate-surface-area	0.0078	0.0029	2.635
3	HOMO-1 energy	1.73	0.40	4.323
4	Max total interaction for a C-H bond	-16.23	5.17	-3.141
5	PNSA-2 Total charge weighted PNSA [Quantum-Chemical PC]	-0.021	0.010	-2.062
	N= 78,R^2=0.76, F=44.92, RMS=0.5307			

Table 2 Descriptors, Coefficients, Standard Error, and t-Test Values for the Linear model

linear model with a high number of terms is not very practical, especially when a high number of experimental descriptors are used, so stepwise variable selection was applied.

The forward stepwise regression procedure consists simply in a step-by-step addition of the best descriptors to the model that leads to the smallest standard deviation (S), until there is no-other variable outside the equation that satisfies the selection criterion. The stepwise regression technique requires much less linear regressions and need not search for optimal variables in the whole data set.

3.2 Theory of radial basis function neural network (RBFNN)

In the research fields of QSAR or QSPR, nonlinear algorithms are usually used. The nonlinear methods can represent the complicate relationships between the activity or property of the molecules and the structures well. And the nonlinear models usually give better results. Artificial neural networks have been found much popularity in the studies. The theory of RBFNN has been extensively presented in some papers [23, 24]. Here only a brief description of the RBFNN principle was given. Figure 1 shows the basic network architecture. It consists of an input layer, a hidden layer, and an output layer.

The input layer does not process the information; it only distributes the input vectors to the hidden layer. The hidden layer of RBFNNs consists of a number of RBF units (n_h) and bias (b_k). Each hidden layer unit represents a single radial basis function, with associated center position and width. Each neuron on the hidden layer employs a radial basis function as a nonlinear transfer function to operate on the input data. The most often used RBF is a Gaussian function that is characterized by a center (c_j) and a width (r_j). The RBF functions, by measuring the Euclidean distance between the input vector (**x**) and the radial basis function center (c_j), performs the nonlinear transformation with RBF in the hidden layer as given below

$$h_j(X) = \exp\left(-\left\|X - c_j\right\|^2 / r_j^2\right) \quad (1)$$

In which h_j is the notation for the output of the j_{th} RBF unit. For the jth RBF, c_j and r_j are the center and the width, respectively. The operation of the output layer is linear, which is given below

$$y_k(X) = \sum_{j=1}^{n_k} w_{kj} h_j(X) + b_k \quad (2)$$

Where y_k is the k_{th} output unit for the input vector **x**, w_{kj} is the weight connection between the k_{th} output unit and the j_{th} hidden layer unit, and b_k is the bias. It can be seen from eqs 1 and 2, designing a RBFNN involves selecting centers, number of hidden layer units, width, and weights.

There are various ways for selecting the centers, such as random subset selection, K-means clustering, orthogonal least squares learning algorithm, RBF-PLS, etc. The widths of the radial basis function networks can either be chosen the same for all the units or can be chosen differently for each unit. In this paper, considerations were limited to the Gaussian functions with a constant width, which was the same for all units. The adjustment of the connection weight between hidden layer and output layer is performed using a least-squares solution after the selection of centers and width of radial basis functions.

The overall performance of RBFNN is evaluated in terms of a root-mean-squared error (RMS) according to the equation below:

$$RMS = \sqrt{\frac{\sum_{i=1}^{n_k} (y_k - \hat{y}_k)^2}{n_k}} \qquad (3)$$

Where y_k is the desired output and $\hat{y}_k$ is the actual output of the network; n_k is the number of compounds in analyzed set.

In a word, the performance of RBFNN is determined by the values of following parameters: The number n_h of radial basis functions, the center c_j and the width r_j of each radial basis function, the connection weight w_{kj} between the j_{th} hidden layer unit and the k_{th} output unit. The centers of RBFNN are determined with the forward subset selection method proposed by Orr [25, 26]. The optimal width was determined by experiments with a number of trials by taking into account the leave-one-out (LOO) cross-validation error. The one that gives a minimum LOO cross-validation error is chosen as the optimal value.

All programs implementing RBFNN were written in M-file based on a MATLAB script for RBFNN. The scripts were run on a personal computer.

4 Results and discussion

4.1 Results of MLR

MLR was used to develop the linear model for the prediction of LogT using training set. To determine the optimum number of descriptors, a variety of subset size (From 1 to 9) was investigated to build the model. When adding another descriptor did not improve the statistical ability of a model significantly, it was determined that the optimum subset size of descriptors had been achieved. Good correlations with the experimental LogT data were selected based on the squared correlation coefficient (R^2), Fisher criterion (F), squared cross-validated correlation coefficient (R_{cv}^2) and standard error (s^2) of the regression.

In present study, we used the best correlation equation with five descriptors for the analysis. A detailed description of the linear model based on compounds in the training set is summarized in Table 2.

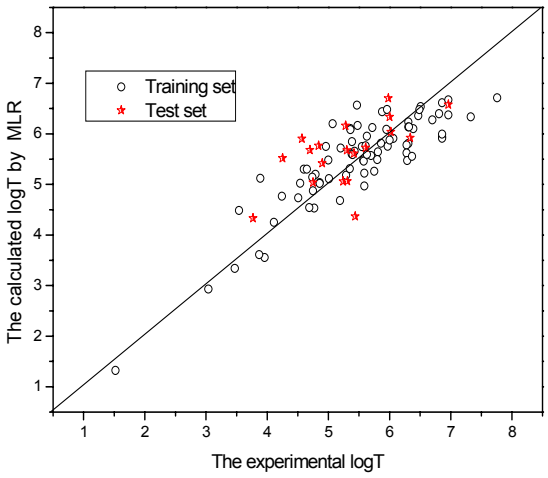

Figure 2. (a) Observed vs. calculated values for Log T by MLR

With the test set, the prediction results were obtained, the statistical parameters were R^2=0.619; F=10.58; and RMS =0.689. The predicted versus observed value based on MLR was shown in Table 1. Figure 2a shows the predicted versus observed LogT for all of the 97 compounds studied, the training set and the test

set. Figure 2b shows the plot of residual vs. calculated Log T.

4.2 Results of RBFNN

After the establishment of a linear model, RBFNN is used to develop a non-linear model based on the same subset of descriptors. To obtain better results, the parameters that influence the

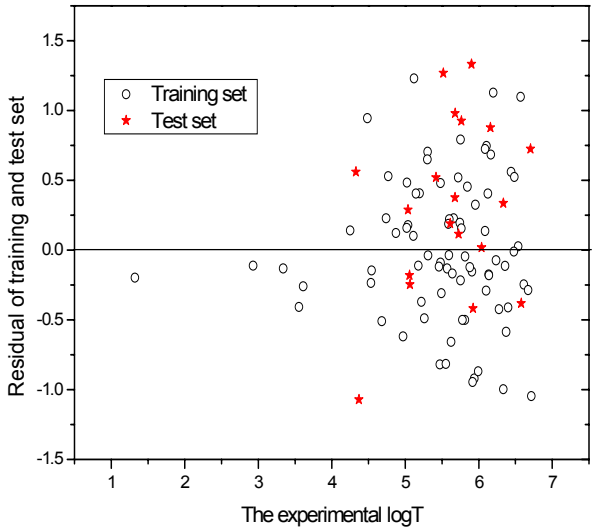

Figure 2 (b) Plot of residual vs. calculated LogT by MLR

performance of RBFNN were optimized. The selection of the optimal width value for RBFNN was performed by systemically changing its value in the training step. The value that gives the best leave one out (LOO) cross-validation result was used in the model.

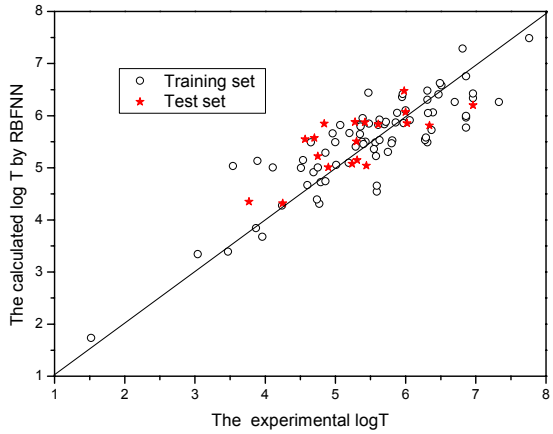

Figure 3 (a) Observed vs. calculated values for LogT by RBFNN

For this data set, the optimal value was determined as 2.50. The corresponding number of centers (hidden layer nodes) of RBFNN is 13. The obtained model had a correlation coefficient R^2 =0.863, F=221.4, with an RMS error of 0.5231 for the training set. The statistical parameters of test set were R^2 =0.76; F=22.7; and RMS=0.5281. The predicted results of the nonlinear models for both training and test set are shown in Table 1. Figure 3a shows the predicted versus observed LogT for all of the 97 compounds studied, the training set and the test set. Figure 3b shows the plot of residual vs. calculated Log T. It gave a better random distribution of the residuals. Comparison of the correlation models obtained with RBFNN and MLR, it is clear that the whole performance of RBFNN is a little better than that obtained by MLR.

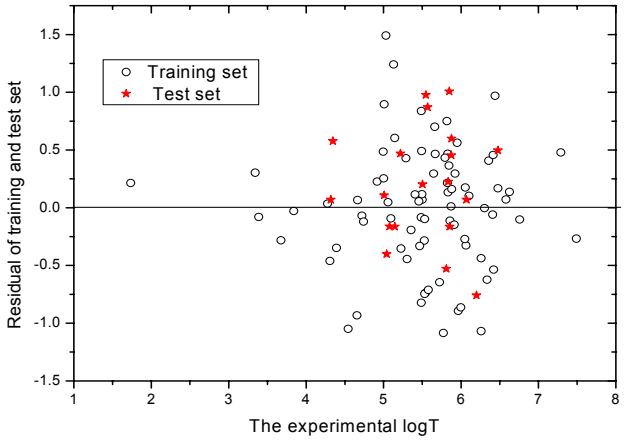

Figure 3(b) Plot of residual vs. calculated LogT by RBFNN

In the present study, principal component analysis (PCA) was also used to study the variation among the chemicals. In PCA, the original data matrix is decomposed into new latent variables and the variation among the objects, here compounds, is illustrated in score plots. As can be seen from the Figure 4, the structures of the compounds are diverse in both sets. The training set with a broad representation of the chemistry space was adequate to ensure models' stability and the diversity of prediction set can prove the predictive capability of the model.

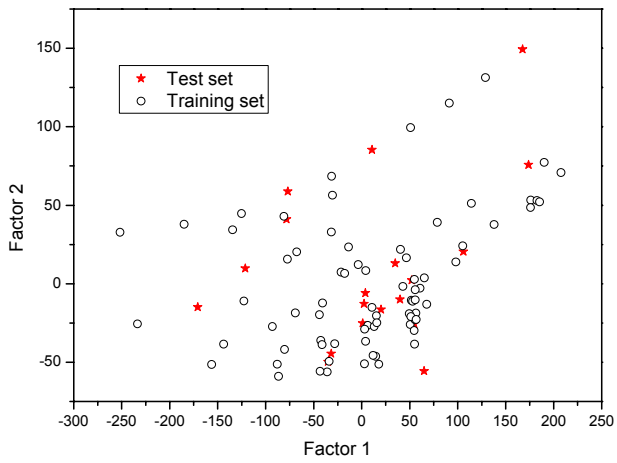

Figure 4 PC1-PC2 score plot

4.3 Discussion of the input parameters

Odor threshold of volatile organic compounds is one of the more commonly used criteria for studying compounds in air. These pungent sensations arise from the activation of receptors are present within the free endings of the trigeminal nerve [27]. Accordingly, chemesthesis is an aspect of the somatic sensory system. Therefore, the factors influencing the properties are complex. They are not only dependent on the characteristic of physiological factors, but also on the physicochemical properties and the molecular structure. From the view of chemistry, property is determined by structure. The aim of this study is to seek the structure factors that influence the odor threshold of the aliphatic alcohols. By interpreting the descriptors in the models, it is possible to gain some insight into structural factors that are likely to relate to the properties studied.

There are five descriptors in the models, which encode different structure feature of each compound. Average complementary information content (order 0) (AIC0) belongs to topological descriptors. It describes the size, shape and branching information of the molecules and gives some information about the hydrodynamic friction factors. Another descriptors, approximate-surface-area (SA), also reflects the size or the shape of the molecular. A molecule with larger SA will be distributed less in air. The presence of this descriptor in our model underlines the importance of the molecular size for the process. They bring a positive contribution to the odor threshold. This observation implies that, all things being equal, increasing the value of the descriptors can lead to the larger values of odor threshold.

HOMO-1 energy denotes the energy of the second highest occupied molecular orbital. HOMO-1 is crucially important in governing molecular reactivity. Molecules with high HOMO-1s are more able to donate their electrons and are hence relatively reactive compared with molecules with low-lying HOMO-1. That is, HOMO-1 may serve as a measure of the excitability of a molecule: all else being equal, the larger the energy, the more easily it will be excited. This can be seen from the positive coefficients in the model.

Max total interaction for a C-H bond (I_{MAX}) is a quantum mechanical energy-related descriptor. This group of descriptors characterizes the total energy of the molecule in different energy scales and the intramolecular energy distribution using different partitioning schemes. As we know, the PNSA-2 Total charge weighted PNSA [Quantum-Chemical PC] (PNSA-2) is one of the charged partial surface area (CPSA) type, which are based on the surface area of the whole molecule and on the charge distribution in the molecule, so they combine shape and electronic information to characterize the molecule, and therefore they encode features responsible for polar interactions between molecules. The above two descriptors have negative regression coefficient which indicate that the LogT is inversely proportional to these descriptors.

5. Conclusion

In this study, QSPR models for odor threshold of 97 aliphatic alcohols were developed using MLR and RBFNN based on some calculated chemical descriptors. Satisfactory results were obtained with the proposed method. Additionally, nonlinear RBFNN model based on the same sets of descriptors showed better predictive ability.

This paper provided a simple and straightforward way to predict LogT of aliphatic alcohols from their structures alone and gave some insight into structural features related to this property of the compounds.

Acknowledgement

The authors thank the National Natural Science Foundation of China (NSFC) Fund (NO. 51073132) for financial support.

References

[1] J. E. Cometto-Mufiiz and W. S. Cain, Ann. N.Y. Acad. Sci., 641 (1992) 137-145.
[2] L. MΦlhave, Ann. N.Y. Acad. Sci., 64 (1992) 46-55.
[3] R. Atkinson, Atmos. Environ., 34 (2000) 2063-2101.
[4] R. G. Derwent, M. E. Jenkin, and S. M. Saunders, Atmos. Environ., 30 (1996) 181–199.
[5] G. Fráter, J. A. Bajgrowicz, and P. Kraft, Fragrance chemistry Tetrahedron., 54 (1998) 7633–7703.
[6] M. Gautschi, J. A. Bajgrowicz, and P. Kraft, Chimia., 55 (2001) 379–387.
[7] P. Laffort, and F. Patte, J. chromatogr., 406 (1987) 51-74.
[8] S. Mihara, and H. Masuda, J. Agric. Food Chem., 36 (1988) 1242-1247.
[9] U. I. Seemai, D. M. Ennis, H. V. Secor, L. Clawson, and J. Palen, Chemical Senses, 14 (1989) 395-405.
[10] B. Winter, QSAR in olfaction. Fauchere.J .L. Ed., QSAR: Quantitative Structure-Activity Relationships in Drug Design. Alan R. Liss, New York, 1989 , p. 401 -405.
[11] P. A. Edwards, and P. C. Jurs, Chemical Senses., 14 (1989) 281-291.
[12] P. A. Edwards, L. S. Anker1 and P. C. Jurs, Chemical Senses., 16 (1991) 447-465.
[13] M. Chastrette, Trends in structure-odor relationship. SAR QSAR Environ. Res., 6 (1997) 215-254.
[14] T. Yamanaka, Chemical Senses, 29 (1995) 471-475.
[15] M. H. Abraham, J. M. R. Gola, J. E. Cometto-Muñiz and W. S. Cain. Chemical Senses, 27 (2002) 95–104.
[16] K. O. Schnabel, H. D. Belitz and C. von Ronson and Z. Lebensm. Unters Forsch., 187(1988) 215-223.
[17] HyperChem 6.01, Hypercube, Inc., 2000.
[18] MOPAC, v.6.0 Quantum Chemistry Program Exchange, Program 455, Indiana University: Bloomington, IN.1999
[19] A. R. Katritzky, V. S. Lobanov and M. Karelson. CODESSA: Training Manual, University of Florida, Gainesville, FL, 1995.
[20] A. R. Katritzky, V. S. Lobanov and M. Karelson. CODESSA: Reference Manual, University of Florida, Gainesville, FL, 1994.
[21] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. Jong, P. J. Lewi and J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics-Part A, Elsevier Science, Amsterdam, 1997.
[22] E. Deconinck, D. Coomans and Y. Vander Heyden, J. Pharm. Biomed, Anal., 43 (2007) 19-130.
[23] Y. H. Xiang, M. C. Liu, X. Y. Zhang, R. S. Zhang, Z. D. Hu, B. T. Fan, J. P. Doucet and A. Panaye, J. Chem. Inf. Comput. Sci., 42 (2002) 592-597.
[24] X. J. Yao, A. Panaye, P. Doucet, R. S. Zhang, H. F. Chen, M. C. Liu, Z. D. Hu, and B. T. Fan, J. Chem. Inf. Comput. Sci., 2004;44:1257-1266.
[25] M. J. L. Orr, Introduction to Radial basis function networks, Centre for Cognitive Science, Edinburgh University, 1996.
[26] M. J. L. Orr, MATLAB routines for subset selection and ridge regression in linear neural networks, Centre for Cognitive Science, Edinburgh University, 1996.
[27] W. L. Silver, T. E. Finger, The trigeminal system. In Smell and Taste in Health and Disease T. V. Getchell, R. L. Doty, L. M. Bartoshuk, and J. B. Snow, Jr., Eds., Raven Press, New York, 1991, p. 97–108.