

기체크로마토그래피에서 QSRR을 통한 PAH 용리시간 예측

金永九

한국가스안전공사 연구개발실
(2001. 4. 4 접수)

Prediction of Gas Chromatographic Retention Times of PAH Using QSRR

Younggu Kim

Division of R & D Korea Gas Safety Corporation, Shilung 429-712, Korea
(Received April 4, 2001)

요 약. 기체 크로마토그래피에서 PAH와 그것의 유도체들의 상대적 용리시간을 인공신경망분석과 다변량 선형 회귀분석을 사용하여 학습한 후, 시험세트들의 상대적 용리시간을 예측하였다. QSRR에서 PAH와 그것의 유도체의 주요한 설명인자는 분자량의 제곱근, 분자의 연결지수(χ), 분자 쌍극자모멘트 및 분자의 길이와 폭의 비율(L/B)이었다. 다변량선형회귀분석에 의하면 큰 분자일수록 용리시간은 길어지며 또한 L/B의 값이 커지면 용리시간이 증가하는 것으로 보아 슬롯이론을 따르고 있음을 알 수 있었다. 반면에 설명인자 사이의 선형 독립성에 영향을 받지 않는 인공신경망 분석결과에 의하면 분자량과 분자 쌍극자 모멘트가 주요한 인자로 작용하고 있었다. 시험세트의 예측 정확도를 나타내는 분산은 선형회귀분석에서는 1.860, 인공신경망분석방법에서 0.206으로서 인공신경망 분석법이 다변량회귀분석보다 더 좋은 예측방법임을 알 수 있었다.

ABSTRACT. Retention relative times(RRTs) of PAH molecules and their derivatives in gas chromatography are trained and predicted in testing sets using a multiple linear regression(MLR) and an artificial neural network(ANN). The main descriptors of PAHs and their derivatives in QSRR are the square root of molecular weight(\sqrt{mw}), molecular connectivity(χ), molecular dipole moment(D) and length-to-breadth ratios(L/B). The results of MLR shows that a heavy molecule has a propensity for long retention time. L/B closely related with slot model is a good descriptor in MLR. On the other hand, ANN which is not effected by the linear dependencies among the descriptors were exclusively based on molecular weight and molecular dipole moment. The variances which shows the accuracy of prediction for retention times in testing sets are 1.860, 0.206 for MLR and ANN, respectively. It was shown that ANN can exceed the MLR in prediction accuracy.

서 론

디젤 차량의 배기가스, 카본블랙과 기타 환경오염원으로부터 배출되고 있는 다방향탄화수소(PAH: Polycyclic Aromatic Hydrocarbon)들은 이산화질소 등의 NOx와 대기 중에서 반응하여 니트로화된 PAH분자를 생성할 수 있으며 이러한 니트로계-PAH분자는 살모넬라 등의 박테리아와 동물세포에 돌연변이 및 인체에 암을 유발

시킬 수 있는 것으로 알려졌다.^{1,2} 대기 중에 존재하는 PAH의 농도는 10 ppb 이하이지만³ 환경 문제를 야기하므로 이에 대한 연구가 절실하다.

이를 위하여 많은 종류의 PAH에 대한 정성·정량 분석이 필요하며 PAH분자들은 분자 구조 및 물리화학적 성질이 유사하여 크로마토그래피의 용리 시간에 두 가지 이상의 피크가 중복될 수 있다. GC에 의한 분석 방법은 LC에 의한 분석보다 분리도 및 감도가 높아 유

용한 장비가 될 수 있으며, 분자의 입체적 구조와 물리화학적 성질은 관련이 깊다는 사실에 의거하여 크로마토그래피상에서 PAH 분자들의 용리 시간을 예측하는 것도 가능하다.^{4,9}

기체크로마토그래피의 용리 시간은 고정상의 불성, 이동상 종류와 유속, 분리하고자 하는 용질의 물리화학적 성질 및 오븐의 온도 등에 의하여 결정된다. 이동상과 유속은 일정하므로, 고정상과 용질의 물성 관계들 기초로 용리 시간을 예측(QSRR: quantitative structure retention time relationship)할 수 있으며 대표적인 QSRR은 다변량회귀분석(MLR: multiple linear regression)과 인공신경망(ANN: artificial neural network)방법이 있다.¹⁰⁻¹¹ 기체크로마토그래피나 액체크로마토그래피에서 QSRR에 관한 기존의 연구는 MLR으로 수행되었으며, 단백질의 2차 구조 예측,¹²⁻¹³ 화합물에 따른 녹성도 예측 등¹⁴ 연구에 ANN 사용을 많이 하지만 크로마토그래피에 대한 QSRR 연구는 상대적으로 적은 상황이다.

각 PAH는 기체크로마토그래피의 오븐의 온도 조건, 즉 일정온도 및 온도 나단계 프로그램 등에 따라 용리 시간이 달라지나 본 연구에서는 PAH 분자들의 분리가 가장 잘 되는 온도 조건에 고정하여 상대적 용리시간(RRT: retention relative time)에 영향을 미치는 각 PAH의 설명인자를 구하여 ANN과 MLR에 각각 적용하여 용리 시간 예측에 관한 정확도를 비교·분석하였다.

실 험

기체크로마토그래피를 통한 PAH와 nitro-PAH에 대한 자료 분석

기체크로마토그래피 분석 조건은 다음과 같았다. Hewlett Packard HP5890 series II plus GC/MSD 5972 (direct inlet), 칼럼은 HP-35(35% diphenyl-65% dimethylpolysiloxane, 30 m 0.32 mm 0.25 μ m), 이동상 기체는 헬륨, 유속은 0.6 ml/min, 오븐 온도 프로그램은 80 $^{\circ}$ C, 10 $^{\circ}$ C/min, 250 $^{\circ}$ C, 10 $^{\circ}$ C/min, 280 $^{\circ}$ C, 3 min이었다. HP7673 auto injector, 주입구 온도는 280 $^{\circ}$ C(무분할), Transfer line의 온도는 300 $^{\circ}$ C, SIM mode(solvent delay: 3.0 min). 이러한 분석조건을 통하여 얻은 실험 자료를 기초로, 불 속에 녹아 있는 각 PAH분자들의 용리 시간을 나프탈렌의 용리 시간(6.2 분)으로 나누어 상대적 용리 시간(RRT)을 계산하였다.

이 론

분자구조의 설계와 모델링

분자구조는 INSIGHT II에서 다음의 과정을 거쳐서 설계하였다. (a) 수소원자가 미포함된 PAH분자를 설계한 후, (b) MOPAC프로그램을 사용하여 각 분자의 최저에너지 구조를 얻어 분자의 기하학적 모양, 분자내의 각 결합 각도, 결합 길이를 구한다.

설명인자(Descriptor)의 설계

설명인자는 다음의 4가지로 분류할 수 있다. (a) 위상학적 설명인자에 조각 설명인자(fragment descriptor), 분자연결지수(molecular connectivity indices),¹⁵ 환경적 설명인자(environment descriptor) 등이 있으며 (b) 기하학적 설명인자에는 주관성 모멘트(principal moments of inertia), 반데르발스 부피, 분자의 면적 및 모양, 탄소원자 전체수, (c) 전기적 설명인자에는 전체 전하량 및 분자 쌍극자 모멘트 (d) 불성 관련 설명인자에는 응고열, 비점, 분자량, 분배상수(n-옥탄올과 물에 대한 분배상수), 굴절률 등이 있다. 위의 각 방법으로 설명인자를 계산한 후 SAS를 사용한 단일회귀분석에서 F값이 4.0보다 작거나 인자분석을 통하여 주요 설명인자와 상관관계가 낮은 설명인자는 제외하였다.⁹

평균구조에 가까운 구조의 PAH 분자 길이와 폭의 비율(L/B), 기체상에서 분자의 확산 속도와 관계하는 분자량의 제곱근(sqmw), 분자내의 결합에 대한 성격을 알려주는 분자연결지수(χ_c), 그리고 고정상과 용질사이에 상호 인력에 영향을 미칠 수 있는 분자 쌍극자 모멘트(D) 등의 설명인자들을 순차적 회귀분석(stepwise regression)을 통하여 얻을 수 있었다.

나프탈렌을 포함한 16개 물질에 대한 설명인자들, 학습세트에 대하여는 Table 2, 시험세트에 대하여는 Table 3에 각각 나타내었다. MLR에서 각 설명인자 사이에 선형독립이 요구되므로 각각의 설명인자사이의 상호의존성 여부를 인자분석으로 계산하였다. sqmw와 χ_c 는 선형적으로 의존성이 낮아(Table 4참조), L/B와 D에 대해

Table 1. The important descriptors in PAH molecules

Descriptors	Physical meaning
sqmw	square root of molecular weight
L/B	length-to-breadth ratio
χ_c	molecular connectivity
D	molecular dipole moment

Table 2. RRTs of input data for PAH molecules in the training sets

Name	RRTs of PAH	sqmw	$^1\chi_c$	D	L/B
Naphthalene	1.000	11.31	3.405	0.009	1.24
9-nitroanthracene	1.451	12.33	5.320	5.226	1.42
Acenaphthalene	2.596	12.41	4.149	0.263	1.08
Phenathene	4.011	13.34	4.815	0.036	1.46
Anthracene	4.061	14.93	4.809	0.007	1.57
Indeno(1,2,3-cd)pyrene	5.212	16.70	7.720	0.424	1.40
2-nitrofluorene	5.333	14.53	5.111	5.749	1.77
Pyrene	5.448	14.21	5.559	0.013	1.27
Benzo(a)anthracene	6.941	15.10	6.220	0.067	1.58
Chrysene	6.988	15.10	6.226	0.005	1.72
3-nitrofluoranthene	7.363	15.72	6.071	6.63	1.43
Benzo(a)pyrene	9.119	15.87	6.976	0.049	1.50

Table 3. RRTs of input data for PAH molecules in the testing sets

Name	RRTs of PAH	Sqmw	$^1\chi_c$	D	L/B
Acenaphthene	2.426	13.34	4.445	0.081	1.06
Fluorene	3.104	12.88	4.612	0.093	1.57
1-nitropyrene	4.011	15.72	6.605	5.448	1.08
Fluoranthene	5.212	14.21	5.565	0.22	1.22
Benzo(k)fluoranthene	8.726	15.87	6.970	0.092	1.48

Table 4. The correlation matrix of descriptors

	Sqmw	L/B	$^1\chi_c$	D
Sqmw	1.000			
L/B	0.233	1.000		
$^1\chi_c$	0.972	0.132	1.000	
D	0.219	0.335	0.0343	1.000

여 선형독립도가 높은 $^1\chi_c$ 를 MLR에서 설명인자로 사용하였으며, 선형독립성에 의하여 영향을 받지 않는 ANN에서 sqmw도 설명인자로 사용하였다.

인공신경망(Artificial neural network) 이론¹⁵

인공신경망은 인체의 신경구조와 유사하여 신호를 전달받는 부위와 받은 신호를 전달해주는 부분으로 나누어진다. 인공신경망 방법중에 통계분석을 위하여 유용한 역전달 신경망(BPN: back propagation neural network)은 자료의 입력 층(input layer), 은닉 층(hidden layer) 및 출력 층(output layer)으로 구분된다(Fig. 1참조).

학습 단계(learning phase)는 역전달 신경망의 은닉 층수를 변화시키면서 출력 층의 결과치와 실험치와의 차를 최소화하는 과정이며, 입력 층의 뉴런 값은 항상 0과 1사이이므로 PAH분석을 위한 설명인자들도 식 (1)

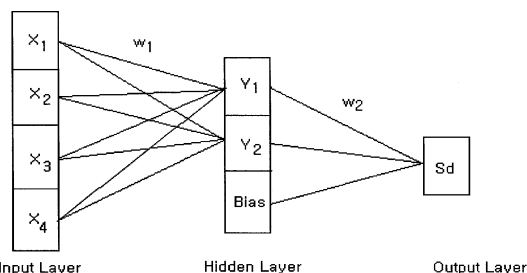


Fig. 1. Three layer back propagation neural network topology in training.

에 의하여 입력 수치를 조절하여야 한다.

$$x'_i = (x_i - x_{\min} - 0.1) / (x_{\max} - x_{\min} + 0.1) \quad (1)$$

여기서 x_{\min} 과 x_{\max} 는 각각 설명인자의 최소값과 최대값이다.

보정된 각각의 값(x'_i)을 가중치(w_j)와 곱하여 합한 값을 y_j 로 하면 $y_j = \sum_i w_{ij} x'_i$ 이다.

$$O_j = \frac{a}{1 - \exp(-b(y_j - \beta))} \quad (2)$$

여기서 a , b , 및 β 는 시그모이드 함수에서 정의하는

매개변수들이다.

숨겨진 층(hidden layer)과 출력 층(output layer)사이의 가중치는 w 라 한다. 식 (2)에서 얻어진 함수 O 와 가중치 w 와의 곱의 합($Bk - \sum_{j=1}^N (O_j \cdot w_{jk})$)이다.

이 값을 시그모이드 함수에 대입하여 계산한 값이 출력 층의 값이다.

$$S_k = \frac{\alpha}{1 + \exp(-b(y, \beta))} \quad (3)$$

ANN에서는 모든 학습세트의 입력자료(example)에 대하여 계산된 출력 결과(S_k)와 실험자료 결과(S_{ik})사이의 차이의 제곱으로 정의되는 오차(E)를 최소화하는 것이 중요하다.

$$E = \sum_{exam} \left[\sum_k (S_k - S_{ik})^2 \right] \quad (4)$$

식 (4)에서는 기호를 단순화하기 위하여 입력자료(example)를 exam으로 간략하게 나타내었다. E 값을 최소화하기 위하여 다음의 미분과정이 포함된 델타 법칙을 사용한다. 델타 법칙은 연결 가중치(weight: w)를 부작위의 값($0 < w_{ij} < 1$)으로 시작하여 학습상을 기지면서(5a, 5b)식의 과정을 거쳐 반복적으로 변화시키는 것이다. 위 첨자 n 과 $(n+1)$ 은 최소화 과정의 연속 횟수를 의미한다. 변화된 가중치(weight)는 다시 BPN 알고리즘에 입력된다.

$$w_{ij}^{n+1} = w_{ij}^n - \eta \frac{\partial E}{\partial w_{ij}} \quad (5a)$$

$$w_k^{n+1} = w_k^n - \eta \frac{\partial E}{\partial w_k} \quad (5b)$$

여기서 w_{ij}^{n+1} 와 w_k^{n+1} 은 변화된 가중치이며, 연속적인 반복과정을 통하여 보정된다.⁶

실험자료(RRT)와 각각의 설명인자를 BPN알고리즘에 입력하여 학습반복과정을 기치면서 가중치는 계속 변화한다. 오차함수인 E 값의 요동이 적고 0에 가까운 값으로 수렴할 때까지 위의 반복과정을 수행한다.

시그모이드 함수(O)의 미분 과정은 식 (6)으로 정의된다.

$$\frac{\Delta O_j}{\Delta \epsilon_j} = \frac{O_j(x + \Delta \epsilon_j) - O_j(x)}{\Delta \epsilon_j} \quad (6)$$

우리 연구에서는: 미분구간 $\Delta \epsilon = 0.5$ 로 할 때 오차(E)

가 0에 가까운 값으로 수렴하였다.

숨겨진 층(hidden layer)의 수는 다음과 같이 결정하였다. 학습세트에서 은의 층의 수가 증가하면 학습세트의 분산은 작아지나 시험세트의 분산은 커지므로 시험세트의 분산이 최적이 되는 경우를 고려하여 은의 층의 단위를 구하였다. 설명인자가 4개일 때 은의 층의 최적 수는 2이었다.

결과 및 고찰

머부분 시간을 예측하는 방법은 설명인자들 사이의 선형독립에 근거한 MLR 방법 및 ANN방법 등이 있다. 기존의 MLR은 SAS 등의 통계프로그램을 사용하여 쉽게 분석이 가능하지만 실제 훈련세트와 시험세트의 예측결과를 비교하면 ANN방법에 의한 것보다 예측에 관한 정확도가 상대적으로 낮은 것을 알 수 있다.¹³ 용리시간과 관계가 깊으며, 상호 독립성 높은 설명인자를 잘 찾을 때에는 쉽게 MLR방법이 유용하나 상호 독립적인 설명인자가 충분치 못하거나 부분적인 선형 독립성을 띤 설명인자인 경우에는 ANN 방법을 사용하면 양호한 예측 결과를 얻을 수 있다.

본 연구에서는 Table 2의 자료를 학습세트(training set)로 사용하여 MLR로 RRT를 예측하여 보았다. MLR에서 4개의 설명인자를 모두 넣어서 분석하면 결정계수(R^2)는 증가하지만, 결정계수만으로는 분석자료로 사용할 수 없으므로 수정결정계수(R_{adj}^2)를 사용한다. 선형 회귀분석에 사용한 설명인자는: $^1\chi_c$, L B이며, RRT와 설명인자 사이의 관계식은 다음과 같다.

$$RRT = -6.478 + 1.316^1\chi_c + 2.887LB, R_{adj}^2 = 0.6685, N = 11, F = 11.08 \quad (7)$$

식 (7)에 의하면 분자의 결합관계에 관한 정보를 나타내는 분자 연결지수에 의한 영향이 가장 컸으며, L B가 1보다 클수록 PAH는 천천히 용리한다. 이는 슬롯모

Table 5. Statistical parameters in linear regression for PAH molecules in training sets

Indexes of statistical parameters: RRTs for PAH molecules	
Intercept	-6.478
$^1\chi_c$	1.316
L B	2.887
F	11.08
R_{adj}^2	0.6685

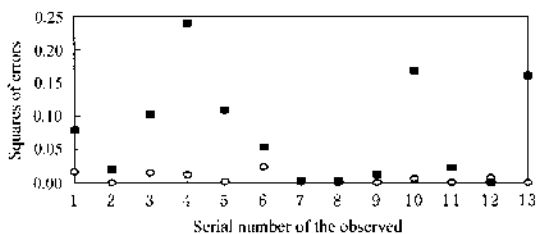


Fig. 2. Plot of results calculated using our BPN program (○) against that using Aoyama's (■) for 13 training sets in carboquinone derivatives where the square of error means $(\log(C)_{\text{observed}} - \log(C)_{\text{calculated}})^2$ and C is the concentration of carboquinone derivatives.

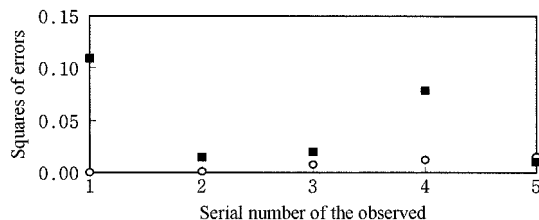


Fig. 3. Plot of results calculated using our BPN program (○) against that using Aoyama's (■) for 5 testing sets in carboquinone derivatives where the square of error means $(\log(C)_{\text{observed}} - \log(C)_{\text{calculated}})^2$ and C is the concentration of carboquinone derivatives.

델로서 잘 설명되며 긴족한 직사각형 모양의 분자가 슬롯사이를 쉽게 왕래하기 때문으로 생각된다.¹⁶

기존 연구의 고속 액체 크로마토그래피에서는 분자 쌍극자모멘트에 의한 영향이 용질의 용리 시간에 영향을 주었으나⁹ 기체크로마토그래피에서는 영향이 적었다. 그 이유는 액체크로마토그래피와는 달리, 기체크로마토그래피에서 PAH분자들은 분리관내의 고정상과 상호작용에 필요한 거리를 취할 수 있는 확률이 작기 때문으로 여겨진다. 상대적 용리시간과 설명인자 사이의 관계식 (7)에 시험세트를 대입하여 계산한 결과는 Fig. 3에 보여주는 바와 같이 정확도가 낮아 분산이 1.861로 큰 값을 보여주었다.

ANN방법으로 분석을 위하여 인공신경망 프로그램을 포트란으로 작성한 후, 우리가 개발한 BPN 프로그램의 정확성을 검증하기 위하여 Aoyama 등이 인용한 분할 자료를 우리 프로그램에 대입하여 Aoyama 연구진의 결과와 비교하여 보았다.^{15,17} Aoyama 등이 사용한 자료 중 13개의 훈련 세트와 5개의 시험 세트를 우리 프로그램으로 분석한 비교 결과를 Table 6과 Fig. 2, 3에 나타내었다. Table 6의 분산 자료에 의하면 우리가 개발한 프로그램이 Aoyama 등이 사용한 프로그램보다, 훈련세트에서 11.5배, 시험세트에서 6.25배의 좋은 결과를 보여주고 있다.

이와 같이 검증된 정확도를 가지는 프로그램에, 각 4개의 설명인자를 입력자료로 사용하여 계산하였다. 이 결과를 Table 7-9와 Fig. 4, 5에 나타내었다. 인공신경망방법(ANN)을 사용하여 PAH분자들의 상대적 용리시간을 분석한 결과는 다음과 같았다. 은의 층과 출력층 사이에 연결된 가중치가 1.735로서 분자연결지수와 1.73값이 증가하면 RRI가 증가하는 경향은 MLR에 의한 결과와 일치하였다. 반면에 분자 쌍극자모멘트가 용리시간에 미치는 영향이 MLR에서는 관찰되지 않았으나 ANN에서는 영향이 있는 것으로 나타났다. Table 6에 의하면 입력층과 은의 층 사이의 연결

Table 6. Comparison of results by neural network of Aoyama's with ours

	Variances	
	Ours	Aoyama's
13 Training sets	0.00649	0.074734
5 Testing sets	0.007398	0.04626

Table 8. The bond weights from the hidden layer to the output layer for BPN in the training sets

Hidden layer number(k)	I	II	bias
W_{2ki}	1.735	-1.440	8.899

Table 7. The bond weights from the input layer to the hidden layer for BPN in the training sets

Bond weight	Descriptors(j)	$\text{sqmw}_{(j-1)}$	$\chi_{(j-2)}$	D	I, B
				$(j-3)$	$(j-4)$
W_{11}		4.312	-3.207	1.774	-1.961
W_{12}		5.77	-4.987	2.713	-2.616

Table 9. Comparison of results between ANN and MLR for the training and testing sets

	Training set		Testing set	
	ANN	MLR	ANN	MLR
Variance	0.2000	1.257	0.110	3.119
The number of datasets(N)	11		5	

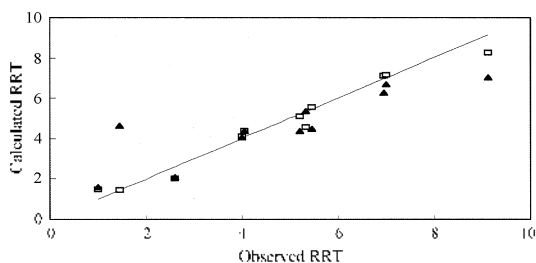


Fig. 4. Plot of calculated RRTs vs. observed RRTs for the PAH molecules in the training set in ANN(□) and MLR(▲) in gas chromatography.

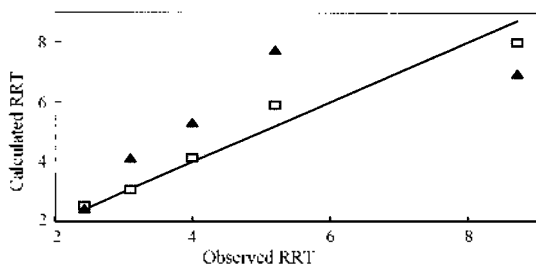


Fig. 5. Plot of calculated RRTs vs. observed RRTs for the PAH molecules in the testing set in ANN(□) and MLR(▲) in gas chromatography.

가중치와 분자 쌍극자 모멘트에서 입력층과 은닉 층 사이의 가중치가 각각 1.774, 2.713이므로 RRT에 큰 영향을 줄 수 있었다. 기존 연구의 HPLC분석에서 분자 쌍극자 모멘트의 값이 커지면 용리 시간은 단축되었으나 기체크로마토그래피에서는 증가된 다른 결과를 보여주었다. 그 이유는 기체크로마토그래피에서 사용한 컬럼은 HPLC용 부극성 컬럼보다 상대적으로 극성도가 큰 중간 극성 컬럼이기 때문으로 생각된다.

결론

본 연구에서는 기체 크로마토그래피에서 Table 2에 보여준 학습세트를 사용하여 PAH 분자 구조에 따

른 상대적 용리 시간에 관하여 고찰하였다. MLR을 사용하여 분자의 설명인자들과 분리관내 용리 시간 사이의 상관관계를 구할 수 있지만 이 방법의 제약 조건은 설명인자사이에 선형독립이어야 한다. 이로 인하여 MLR의 경우, 선형 독립적인 설명인자만으로 QSRR을 수행하는데 불완전하여, 학습세트의 분산이 1.366, 시험세트에서는 1.861이었다. 반면에, ANN을 사용하면 이런 단점을 극복할 수 있어서 학습세트에서 분산은 0.193, 시험세트에서는 0.206 이었다. 결론적으로 ANN은 MLR과 달리 선형적으로 비독립적인 설명인자도 사용하므로 더 좋은 결과를 보여 주었다.

인용문헌

1. Cole, J.; Arlett, C. F.; Lowe, J.; Bridges, B. A. *Mutat. Res.* **1982**, *93*, 213.
2. 한국대기보존학회측정분과위원회. *대기환경과 휘발성유기화합물집*. 1998.
3. Ohgaki, H.; Matsukura, N.; Morino, K.; Kawachi, T.; Sugimura, T.; Morita, K.; Tokiwa, H.; Hirota, T. *Cancer Lett.* (Schanon, Irel.) **1982**, *93*, 213.
4. Johnson, K. F. *The Hammett Equation*; Cambridge University Press: New York U.S.A. 1980.; p213.
5. Kim, Y. G.; Kim W. H.; Pak, H. S. *J. Korean Chem. Soc.* **1998**, *42*(6), 646.
6. Andrea, T. A.; Kalayeh, H. *J. Med. Chem.* **1991**, *34*, 2824.
7. Bartlett, A.; Dearden, J. C. *Quant. Struct.-Act. Related.* **1995**, *14*, 258.
8. Sander, L. C.; Wise, S.A. *Anal. Chem.* **1989**, *61*, 1749.
9. Miram, F. S.; Franco D. M.; Vilma E. F. H.; Rosendo A. Y. *J. Braz. Chem. Soc.* **1999**, *3*, 189.
10. Kim, Y. G. *J. Korean Chem. Soc.* **2000**, *44*(2), 102.
11. Gorgenyi M.; Langenhove H.; Kiraly Z. *J. of Chromatogr. A.* **1995**, *693*, 181.
12. Holly, L. H.; Karplus M. *Proc. Natl. Acad. Sci.* **1989**, *86*, 152.
13. Millik, M.; Kolinski, A.; Skolick, J. *Protein. Engng.* **1995**, *8*(3), 225.

14. Aoyama, T.; Suzuki, Y.; Ichikawa, H. *J. Med. Chem.* **1990**, *33*, 2583.
15. Kier, L. B.; Hall, L. H. *Molecular connectivity in chemistry and Drug Research*; Academic Press : New York, U. S. A., 1986.
16. Sander, L. C.; Wise, S. A. *Anal. Chem.* **1987**, *59*, 2309.
17. Chou, A. H.; Kahn, A. H.; Driscoll, J. S. *J. Med. Chem.* **1976**, *19*, 1302.
-