

고속액체크로마토그래피에서 PAH 분자의 구조에 따른 용리시간 예측

金永九

한국가스안전공사 연구개발실

(1999. 9. 6 접수)

Prediction of Retention Time for PAH Molecule in HPLC

Younggu Kim

Division of R & D Korea Gas Safety Corporation, Shihung 429-712, Korea

(Received September 6, 1999)

요약. 고속액체크로마토그래피에서 PAH분자들의 상대적 용리시간을 다변량선형회귀분석과 인공신경망분석방법을 사용하여 학습시킨 후, 시험 세트의 상대적 용리시간을 예측하였다. PAH의 QSRR에서 주요한 설명인자는 분자연결지수($^1\chi_n$, $^2\chi_n$), 길이와 폭의 비율(L/B) 및 분자 쌍극자 모멘트(D)이었다. 슬롯 모델과 관계있는 L/B는 인공신경망분석방법에서는 적절한 설명인자로 작용하나, 다변량회귀분석에서는 그러하지 못하다. 시험세트에서 용리시간 예측도를 나타내주는 분산은 각각 인공신경망분석방법에서 0.0099, 다변량회귀분석방법에서 0.0114이었다. 인공신경망분석방법이 다변량회귀분석보다 더 좋은 결과를 보여준다.

ABSTRACT. Relative retention times (RRTs) of PAH molecules in HPLC are trained and predicted in testing sets using a multiple linear regression (MLR) and an artificial neural network (ANN). The main descriptors in QSRR are molecular connectivity ($^1\chi_n$, $^2\chi_n$), the length-to-breadth ratios (L/B), and molecular dipole moment(D). L/B which is related with slot model is a good descriptor in ANN, but isn't in MLR. Variances which show the accuracy of prediction for retention times in testing sets are 0.0099, 0.0114 for ANN and MLR, respectively. It was shown that ANN can exceed the MLR in prediction accuracy.

서론

산업이 발달하고 문화 수준이 높아지면서 자동차의 사용량은 급격히 증가하여 특히 도심지 및 교통 정체 구역 등에는 자동차로 인한 배기가스 문제가 심각하다. 배기가스 중의 독성 물질들인 SOx, NOx 및 휘발성 유기화합물(VOC) 등은 인체에 치명적인 영향을 줄 수 있으므로 이에 대한 근본적인 문제 해결은 매우 중요하다. 이 중에 벤젠고리를 포함하고 있는 방향족탄화수소를 다방향족탄화수소(Poly Aromatic Hydrocarbon: PAH)라하는 데 이러한 물질이 공기 중의 이산화질소 등과 결합하면 돌연변이, 암을 유발할 수 있다는 보고가 있다.^{1,7}

환경적으로 문제가 되고 있는 PAH에 대한 연구를 위하여는 대기 중에 분포하고 있는 각각의 PAH에 대

한 정성 및 정량분석이 선행되어야 한다. 그러나 수십 종에 이르는 PAH 및 이의 유도체사이에는 구조 및 물성이 유사하여 분리·분석이 난해하나, 화합물의 구조와 물성 사이의 상관관계가 높다는 사실을 이용하여 이런 난점을 해결할 수 있다. 이와 관련된 이론은 Hammett의 이론에서 찾아 볼 수 있다.⁸ Hammett의 이론은, 벤조산의 유도체들의 pKa값이 유도체의 종류에 따라 변화한다는 것에서 출발하고 있다. 이러한 화합물의 구조와 성질에 관한 연구는, 원하는 성질을 가지는 유기화합물 합성 등의 여러 분야에 응용되고 있으며, 생명과학분야에서는 분자구조와 활성도에 관한 상호관계(QSAR)에서 많이 연구 및 응용하고 있다.^{9,14}

QSAR와 유사하게 용질 분자(PAH)구조와 분리판내에서 용리시간과의 상호관계(QSRR)를 연구하면 쉽게

그 외의 PAH 분자의 용리시간 예측이 가능하다.^{15,16} 용리시간을 예측할 수 있는 방법에는 다변량 회귀분석 방법(multiple linear regression: MLR)과 인공신경망 구조방법(artificial neural network: ANN)이 있으며, 본 연구에서는 고속액체크로마토그래피(HPLC)의 상대적용리시간(relative retention time: RRT)에 영향을 주는 PAH 분자 각각의 설명인자를 구하여 M.L.R.과 ANN의 방법으로 비교 분석하였으며 이중 정확도가 높은 방법으로 PAH의 용리시간을 예측하고자 하였다.

실 험

HPLC를 통한 PAH와 nitro-PAH에 대한 자료의 분석. HPLC 분석 조건은 다음과 같았다. 초기 이동상은 아세토니트릴 수용액 50%이며, 이동상 농도를 1.0 %/min의 속도로 증가시켜서 70%에 도달할 때까지 용리시킨다. 크로마토그래피의 고정상은 ODS Hypersil: 100 mm×4.6 mm I.D., 5 μm 입자이며 분리관의 온도는 40°C이었다. 이러한 분석 조건의 실험 결과로부터, 각 PAH 분자들의 용리시간을 나프탈렌의 용리시간으로 나누어 각 PAH 분자들에 대한 상대적 용리시간(RRT)를 계산하였다.

이 론

분자구조의 설계와 모델링. 분자구조는 INSIGHT

II에서 다음의 두 과정을 거쳐 설계하였다. (a) 그래픽 터미널로 수소 원자가 미포함된 분자들을 스케치한 후 저장한다. (b) MOPAC 프로그램을 통하여 각 분자에 대한 에너지를 최소화한다. 이 과정을 통하여 기하학적인 모양, 분자의 결합 각도, 결합 길이를 구할 수 있다.

설명인자의 설계. 설명인자는 다음의 4 가지 범주로 구분할 수 있다. 첫째, 위상학적인 설명인자에는 조각 설명인자(fragment descriptors), 분자연결지수(molecular connectivity indices), 환경적 설명인자(environment descriptors)등이 있다. 둘째, 기하학적인 설명인자에는 주관성 능률(principal moments of inertia), 반델스발스부피, 면적 및 모양, 탄소원자 수이며, 셋째, 전기적인 설명인자에는 전체 σ전하량(total σ charges) 및 분자 쌍극자 모멘트 등이 있다. 넷째, 물리화학적 설명인자는 응고점, 비점, 분자량, 분배상수(옥탄올과 물에 대한 분배상수), 굴절률 등이 있다.¹⁷ 위의 방법으로 각각의 설명인자를 계산한 후, 일차적으로 MLR을 통하여 회귀에 기여도가 작다고 판정되거나 인자분

Table 1. Descriptors used in PAH molecules

descriptors	physical meaning
L/B	length-to-breadth ratio
¹ χ _v	molecular connectivity
² χ _v	molecular connectivity
D	molecular dipole moment

Table 2. RRTs and the input data for PAH molecules in the training set, where the solvent is gradient acetonitrile (50/70/1)

name	RRTs of PAH	L/B	1	2	D
naphthalene	1.000	1.24	3.405	2.732	0.009
acenaphthalene	1.175	1.08	4.149	2.584	0.263
2-nitrofluorene	1.317	1.77	5.111	3.932	5.749
acenaphthene	1.518	1.06	4.445	3.226	0.081
fluorene	1.562	1.57	4.612	3.448	0.093
phenathene	1.684	1.46	4.815	3.363	0.036
1-nitropyrene	1.938	1.08	6.605	6.605	5.448
3-nitrofluoranthene	1.998	1.43	6.071	6.071	6.63
fluoranthene	2.099	1.22	5.565	4.399	0.22
chrysene	2.709	1.72	6.226	4.335	0.005
benzo(a)anthracene	2.781	1.58	6.220	4.335	0.067
benzo(b)fluoranthene	3.317	1.40	6.976	5.150	0.286
benzo(a)pyrene	3.463	1.50	6.976	5.123	0.049
dibenz(a, h)anthracene	3.947	1.79	7.637	5.579	0.067
benzo(g, h, i)perylene	4.096	1.12	7.720	5.742	0.097

Table 3. RRTs and the input data for PAH molecules in the testing set, where the solvent is gradient acetonitrile (50/70/1)

name	relative retention time of PAH	L/B	1	2	D
9-nitroanthracene	1.518	1.42	5.320	3.925	5.226
anthracene	1.803	1.57	4.809	3.380	0.007
pyrene	2.208	1.27	5.559	4.123	0.013
benzo(k)fluoranthene	3.402	1.48	6.970	5.166	0.092
indeno(1,2,3-cd)pyrene	4.096	1.40	7.720	5.723	0.424

Table 4. The correlation matrix of descriptors

	L/B	${}^1\chi_v$	${}^2\chi_v$	D
L/B	1.000			
${}^1\chi_v$	0.2361	1.000		
${}^2\chi_v$	0.1777	0.8467	1.000	
D	0.3091	0.0946	0.3337	1.000

석을 통하여 주요 설명인자와 상관관계가 높은 설명인자는 제외하였고 설명인자 모두를 순차적 회귀(stepwise regression) 분석으로 다음의 4개의 주요한 설명인자 (L/B, ${}^1\chi_v$, ${}^2\chi_v$, D)를 얻을 수 있었다. 여기서 이들의 물리적 의미는 Table 1에 설명한 바와 같이 L/B는 PAH분자들과 같이 평면구조를 가지는 화합물 분자의 길이와 폭의 비율, ${}^1\chi_v$, ${}^2\chi_v$ 는 분자연결지수로서 분자결합에 대한 성격, D는 분자쌍극자모멘트이다. 나프탈렌의 9개 PAH 분자들에 대한 상대적 용리시간과 설명인자들을 학습세트에 대하여는 Table 2, 시험세트에 대하여는 Table 3에 각각 나타내었다.

각 설명인자 사이에 선형독립성이 요구되는 선형회귀분석에서는, 4개의 설명인자 사이의 상호 선형 독립성을 계산하였다. 그 결과를 Table 4에 나타내었으며 ${}^1\chi_v$ 와 ${}^2\chi_v$ 는 높은 상호의존성(상관계수는 0.8467)을 보여 주었다. 그리하여 ${}^1\chi_v$ 의 L/B, D에 대한 각각의 상관계수는 0.2361, 0.0946 이며, ${}^2\chi_v$ 의 경우에는 0.1777, 0.3337 이어서, 타 설명인자에 비하여 선형독립성이 높은 ${}^1\chi_v$ 를 분자연결지수의 설명인자로 선택하였다. 그리하여 다변량회귀분석에서 L/B, ${}^1\chi_v$, D 등의 설명인자를 학습세트에 사용하였고 독립성이 낮은 ${}^2\chi_v$ 도 인공신경망분석방법에서 설명인자로 사용하였다.

분자 연결지수의 결정 ${}^m\chi_v$ 에서(m은 급수를 표시하고 v는 원자가를 나타낸다), 분자의 골격구조로 표시된 원자의 원자가전자수에서 그 원자에 연결된 수소원자의 수를 빼 값을 δ 로 표시하였다. 첫째 원자가의 차수(v)은 인접한 원자 i와 j로부터 값이 얻어진다. 예로서 1-

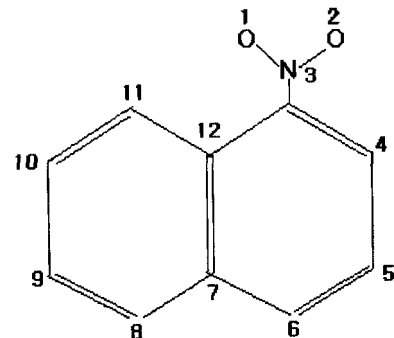


Fig. 1. Structure of 1-nitronaphthalene. The values of delta are calculated for hetero atoms.

니트로나프탈렌의 ${}^1\chi_v$ 를 계산하면 다음과 같이 된다.¹⁸

$${}^1\chi_v = \sum_{q=1}^L (\delta_i \delta_j) q^{(-1/2)}$$

식에 의하여 아래와 같이 표시될 수 있다.

$$\begin{aligned} {}^1\chi_v &= \sqrt{\frac{1}{6 \cdot 5}} + \sqrt{\frac{1}{6 \cdot 5}} + \sqrt{\frac{1}{5 \cdot 4}} + \sqrt{\frac{1}{4 \cdot 3}} + \sqrt{\frac{1}{3 \cdot 3}} + \sqrt{\frac{1}{3 \cdot 3}} \\ &+ \sqrt{\frac{1}{3 \cdot 4}} + \sqrt{\frac{1}{4 \cdot 3}} + \sqrt{\frac{1}{3 \cdot 3}} + \sqrt{\frac{1}{3 \cdot 3}} + \sqrt{\frac{1}{3 \cdot 3}} + \sqrt{\frac{1}{3 \cdot 4}} \\ &+ \sqrt{\frac{1}{4 \cdot 4}} = 3.660 \end{aligned}$$

역 전달신경망(Back propagation neural network : BPN) 이론. 인공신경망방법은 두뇌의 신경구조와 유사하여 전달부위와 전달받는 부위가 있다. Hopfield 등은 역전달신경망구조를 개발하였다.¹⁹ BPN은 입력층, 숨겨진 층 및 출력층으로 나누어지며 Fig. 2에서 각 층의 신경단위(neuron) 또는 마디(node)는 결합에 의하여 연결되어 있다. 각 입력층 마디는 독립적이나, 출력층 마디는 다른 종속적 변수이다. 입력 층 및 숨겨진 층은 각각 1.0에 해당하는 값의 바이어스(bias)를 첨가할 수 있다. Fig. 2에서 입력 층은 4 단위, 숨겨

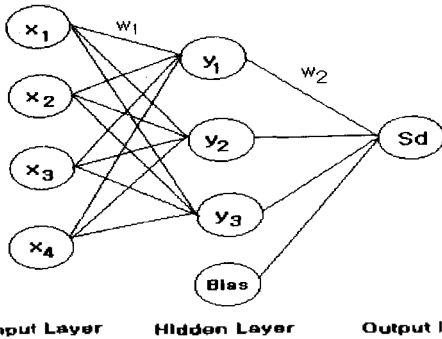


Fig. 2. Three layer back propagation neural network topology in training.

진 층(Hidden Layer)는 3 단위, 출력 층은 1 단위로 하는 인공신경망을 나타내주고 있다. 역전달신경망은 다음의 학습상(learning phase)을 보정할 때 사용된다.^{20,21}

숨겨진 층의 y_j 값은 $y_j = \sum w_{ij}x_i$ 으로 계산가능하며 여기서 w_{ij} 는 입력 단위 i 와 숨겨진 단위 j 사이에 연결되는 결합 가중치이다. 출력치 O_j 는 종속변수 y_j 에 대한 함수로서 (1) 식으로 정의된다.

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

여기서 a , b 및 β 는 시그모이드 함수를 정의하는 계수이며, x_i 는 입력되는 설명인자이다. $S_k = \sum_{j=1}^n W_{jk} \cdot O_j$ 로 정의되며, W_{jk} 는 출력 단위 k 와 숨겨진 층 j 사이에 연결된 결합가중치이다. 신경망으로 계산된 O_j 는 0에서 1사이의 값을 가진다. 학습상(training phase)에서는 학습 세트의 모든 설명인자를 인공신경망구조에 대입하여 계산된 출력치(S_i)와 실험 결과(S_d)사이의 오차의 제곱 합을, (2)식의 E 로 정의한다.

$$E = \sum_k (S_i - S_d_k)^2 \quad (2)$$

E 값을 최소화하기 위하여 다음의 역전달신경망을 수행한다. 오차함수 E 는 결합가중치에 의하여 변화하며, 이 값을 최소화하기 위하여 결합가중치에 대하여 오차 함수를 미분한다. 이와 같이 변화된 결합가중치(hond weights)는 (3)식으로 표현된다.

$$W_{ij}^{n+1} = W_{ij}^n - \eta \frac{\partial E}{\partial W_{ij}} \quad (3)$$

오차함수 E 를 최소화하기 위하여 가장 공동적인 질

차는 델타 법칙에 따른다.²² 델타 법칙은 연결가중치를 초기에 무작위로 정하여진 작은 값에서부터 시작하여 반복적으로 변화시키는 것이다. 위 첨자 n 과 $n+1$ 은 최소화 연속과정의 반복회수를 나타낸다. η 는 일반적으로 1보다 훨씬 작은 값을 지니는 학습률(learning rate)이다. 학습 세트에서 (3)식과 같이 변화된 결합가중치를 역전달신경망에 넣어, 계산된 출력치가 실험 결과에 수렴할 때까지 수 많은 반복과정을 수행한다.

신경망구조에 입력(x) 또는 출력 자료(S_d)는 0과 1 사이로 크기가 조절되어야 하므로 크기 조절식(scaling equation) (4) 식을 사용하여 자료를 정리한다.

$$e_{ij}^1 = (e_{ij} - e_{min} + 0.1) / (e_{max} - e_{min} + 0.1) \quad (4)$$

여기서 e_{max} , e_{min} 은 각각 최소 및 최대값이다.

자료를 BPN 알고리즘에 입력한 후, 학습 반복과정을 통하여 학습 값들의 가중치(weights)는 변화한다. 오차함수 E 값의 요동이 적고, 지역적인 극소치에서 벗어나 신속한 수렴을 위하여 학습률 및 시그모이드의 미분항에서 제한사항이 필요하다. 일반적으로 이러한 제한사항은 역전달신경망(BPN)을 수행하여 실험적으로 얻어지며, 우리 분석 자료에서 학습률과 미분구간 ΔE 의 제한사항은 다음과 같이 구할 수 있었다. 첫째, 학습률(learning rate: η)은 $\begin{cases} 0.055 & \text{for } C_1 > 0.122 \\ 0.022 & \text{for } C_1 > 0.122 \end{cases}$

여기서 C_1 은 E 값에 해당한다. 둘째, (3) 식에서 시그모이드 함수(O_j)의 미분과정은 (5) 식으로 표현된다.

$$\frac{\Delta O_j}{\Delta E} = \frac{O_j(x + \Delta E) - O_j(x)}{\Delta E} \quad (5)$$

미분 구간 ΔE 은 다음의 조건으로 제한된다.

$$\begin{cases} 5.0 & \text{for } C_1 > 0.122 \\ 0.050 & \text{for } C_1 > 0.122 \end{cases}$$

숨겨진 층(Hidden layer)의 수는 다음과 같이 결정된다. 학습세트에서 숨겨진 단위의 수가 증가하면 일정 크기까지는 분산이 작아지나, 시험세트에 대한 분산은 그러하지 못하기 때문에 시험세트에 대한 분산이 가장 작을 때를 숨겨진 단위의 최적수로 정한다. 본 계산의 경우에 설명인자의 수가 4일 때, 숨겨진 단위 수는 3으로 할 때 시험세트의 분산이 최소치를 가져 숨겨진 단위 수는 3으로 정하였다.

결과 및 고찰

학습세트로 사용한 Table 3 자료를 통계프로그램 SAS으로 MLR분석을 실행하였으며, 그 결과를 Table 5에 나타내었다. 선형회귀분석에서는 일반적으로 설명변수(여기서는 설명인자)의 수가 증가하면 상관계수의 제곱(R²)은 항상 증가하므로, 설명변수의 개수가 다른 모형에 대하여는 R²값의 비교만으로는 불충분하다. 이를 보완하기 위하여 수정결정계수(R²_{adj})를 사용한다. Table 5의 자료는 다음 (6)식으로 표현되며, 각 설명인자의 세트를 (6)식에 대입하여 예측한 결과와 실험적으로 측정된 결과에 대한 상관관계를 Fig. 3에 보여 주었다.

$$RRT = -1.824 + 0.7474^1\chi_1 - 0.121D R^2_{adj} = 0.982, \quad (6)$$

N=15, F=406.0

다변량선형회귀분석의 결과에 의하면 PAH분자들의 용리시간은 분자의 연결지수와 쌍극자 모멘트에 의하여

Table 5. Statistical parameters in multiple linear regression for PAH molecules in the training set, where the solvent is gradient acetonitrile (50/70/1)

indexes of statistical parameters	RRTs for PAH molecules
Intercept	-1.824
χ_1	0.747
D	-0.121
F	406.0
R ²	0.985
R ² _{adj}	0.982

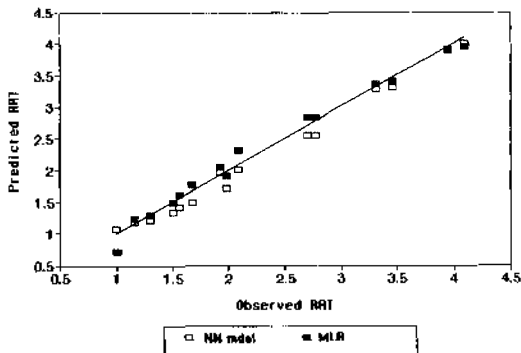


Fig. 3. Plot of predicted RRTs vs. observed RRTs for the PAH molecules in the training set in ANN and M.L.R, where the solvent is gradient acetonitrile (50/70/1) in HPLC.

결정되며 쌍극자 모멘트가 증가하면 용리시간은 감소한다. 즉 극성도가 높은 분자들은 무극성이 높은 분리관의 고정상과의 상호작용이 작으므로 용리 시간이 작다. 또한 분자량과 상관관계가 높은 분자연결지수(χ_1)가 증가하면 분리관과 용질분자의 상호작용에 의하여 친척히 용리됨을 볼 수 있다.

인공신경망은 포트란 프로그램으로 직접 작성한 후, 다변량회귀분석에 사용한 학습세트들을 대입하여 실행하였다. 분석 결과를 Table 6, 7에 나타내었다. Fig. 2의 인공신경망구조와 관련지어 Table 6, 7 결과를 설명하면 다음과 같다. 출력치(Sd)에 영향을 가장 크게 주는 w_{2k1} (숨겨진 층에서 출력층 사이에 존재하는 가중치)에서 절대치가 가장 큰 것이 영향을 많이 주며, 본 연구에서는 w_{231} 이 가장 중요한 영향을 미친다. 숨겨진 층의 y_j 와 연결된 L/B의 w_{1j1} , χ_1 의 w_{121} , χ_2 의 w_{131} , D의 w_{141} 값들의 크기를 고찰하면 각 설명인자들과 상대적 용리상수사이의 상관관계를 알 수 있다. 분자연결지수와 관계된 가중치는 양수 값을 가짐으로 분자량이 커지면 늦게 용출됨을 알 수 있고, 분자 쌍극자 모멘트 즉 분자의 극성도가 커지면 빨리 용출됨을 알 수 있다. Sd 값과 상관된 가중치들을 비교하면 χ_1 의 경우가 가장 크며($w_{131}=0.4349$) 쌍극자모멘트에 대한 가중치의 절대값이 큰 음수($w_{111}=-0.0909$)를 가지는 것으로 보아 쌍극자모멘트가 커지면 RRT값이 작아진다. 이것은 선형회귀분석의 결과와 일치한다. L/B에서 y_j 와 연결된 가중치들은 $w_{111}=0.132$, $w_{231}=0.2851$ 로서, 분자 모양이 정사각형인 L/B=1.0 이면 RRT가 작아지며, 길쭉한(L/B>1.0) 분자의 RRT는 증가한다. 이것은 L/B에 의한 용리시간의 변화를 설명하는 슬롯 모델과

Table 6. The bond weights from the input layer to the hidden layer for BPN in the training set

descriptor(j) bond weight	L/B(j=1)	$\chi_1(j=2)$	$\chi_2(j=3)$	D(j=4)
	w_{1j1}	-0.11675	0.6331	0.2131
w_{1j2}	0.132	0.7268	0.1821	-0.2467
w_{1j3}	-0.03617	0.4349	0.0127	-0.0909

Table 7. The bond weights from the hidden layer to the output layer for BPN in the training set

hidden layer number(k)	I	II	III	bias
w_{2k1}	-0.02475	0.2851	0.7164	0.1119

유사한 결과를 보여준다.²³ 선형회귀분석에서는 L/B의 효과를 단정적으로 결정짓기가 쉬지 않으나 ANN에서는 슬롯 모델을 충분히 설명할 수 있었다. 회귀분석에서는 각 설명인자들 사이에 선형독립성이 낮은 것은 입력 자료로서 배제되나, ANN에서는 선형독립성이 낮은 것, 등의 자료에 대하여도 사용가능하다.

결 론

본 연구에서는 Table 2에 나타난 PAH 분자를 학습 세트(training set)로 하여 QSRR을 수행하였다. 일반적으로 회귀분석을 통하여 QSAR 연구를 쉽게 할 수 있지만 좋은 설명인자를 구한다 하더라도 각각의 설명인자 사이에는 선형독립성이 낮은 경우도 많이 존재한다. 이런 경우에는 QSRR 연구에서 R²값이 작아 미지 분자에 대한 물성 또는 본 연구에서 용리시간 예측이 정확하지 못할 때가 있다. 그러나 선형독립성이 낮아도 역선달신경망을 사용하면 예측 정확도를 높일 수 있다. 본 연구에서도 Fig. 4와 Table 8에 보여주는 바와 같이 분산은 ANN의 경우 0.0099, MLR의 경우 0.0114로서, L/B외 3개의 설명인자를 넣어 계산한 ANN 방법이 MLR에 비하여 높은 적합성을 나타내었으며 평

면형 구조를 띠는 분자에 대하여 슬롯 모델을 잘 설명하고 있다. 결론적으로 비선형 독립성의 설명인자를 배제하는 다변량회귀분석보다, 설명인자들 사이의 독립성과 무관하게 통계분석가능한 인공신경망방법이 QSSR연구에 좋은 결과를 보여 주었다.

인 용 문 헌

1. Cole, J.; Arlett, C. F.; Lowe, J.; Bridges, B. A. *Mutat. Res.* **1982**, *93*, 213.
2. Harely, R. A.; Class, G. R. *Atmospheric Environment* **1995**, *29*(8), 905.
3. 한국대기보전학회 측정분석분과위원회. *대기환경과 휘발성 유기화합물질*. **1998**.
4. Mermelstein, R.; Kiriazides, D. K.; Butler, M.; Mccoy, E. C.; Rosenkranz, H. S. *Mutat. Res.* **1981**, *89*, 187.
5. Ohgaki, H.; Matsukura, N.; Morino, K.; Kawachi, T.; Sugimura, T.; Morita, K.; Tokiwa, H.; Hirota, T. *Cancer Lett.*(Shanon, Irel.) **1982**, *93*, 213.
6. Schuetzle, D.; F. S. C. L.; Prater, T. J.; Tejada, S. B. *Proceedings of the 10th Annual Symposium on Analytical Chemistry of Polluants*; Gordon and Breach Science Publishers: New York, 1980; p193.
7. Pitts, J. N. Jr.; Lokensagard, D. M.; Harger, W.; Fisher T. S.; Mcjia, V.; Shuler, J. J.; Scorziell, G. M.; Katzenstein, Y. A. *Mutat. Res.* **1982**, *103*, 241.
8. Johson, K. F. *The Hammet Equation*; Cambridge University Press: New York, U.S.A. 1975.
9. Andrea, T. A.; Kalayeh, H. *J. Med. Chem.* **1991**, *34*, 2824.
10. Cronin, M. T. D.; Dearden, J. C. *Quant. Struct-Act. Related.* **1995**, *14*, 1.
11. Cronin, M. T. D.; Dearden, J. C. *Quant. Struct-Act. Related.* **1995**, *14*, 117.
12. Bartlett, A.; Dearden, J. C. *Quant. Struct-Act. Related.* **1995**, *14*, 258.
13. Aoyama, T.; Suzuki, Y.; Ichikawa, H. *J. Med. Chem.* **1990**, *33*, 2583.
14. Bisset, L. R.; Fierz, W. J. *Molecular Recognition* **1993**, *6*, 41.
15. Kim, Y. G.; Kim, W. H.; Pak, H. S. *J. Korean. Chem. Soc.* **1998**, *42*(6), 646.
16. Peterson, K. L. *Anal. Chem.* **1992**, *64*, 379.
17. Egolf, L. M.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 616.
18. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
19. Hopfield, J. J. *Proc. Nat. Acad. Sci.* **1984**, *81*, 3088.

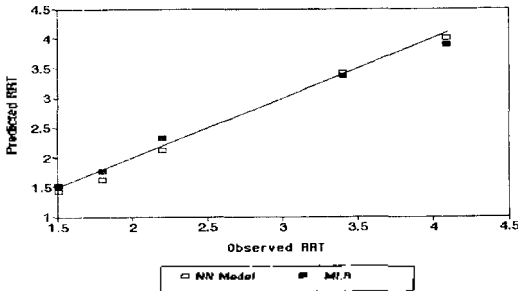


Fig. 4. Plot of predicted RRTs vs. observed RRTs for the PAH molecules in the testing set in ANN and M.L.R, where the solvent is gradient acetonitrile (50/70/1) in HPLC.

Table 8. Comparison of results between ANN and MLR for the training and testing sets

	training set		testing set	
	NN	MLR	NN	MLR
variance	0.0047	0.01422	0.0099	0.0114
the number of datasets (N)	15		5	

20. Devillers, J.; Karcher, W. J. *Applied Multivariate Analysis in SAR and Environmental Studies*; ESCS, EEC, EAEC. Brussels and Luxembourg, 1991; p479.
 21. Schüüman, G.; Mütter, E. *Environmental Toxicology and Chemistry* **1994**, *13*(5), 743.
 22. Rumelhart, D. E.; McClelland, J. L. *Parallel Distributed Processing; Exploration in the Microstructure of Cognition*; MIT Press: Cambridge, 1986; Vols I and II.
 23. Sander, L. C.; Wise, S. A. *Anal. Chem.* **1989**, *61*, 1749.
-