

## 신경회로망을 이용한 Joback Group Contribution Method의 개선

이건홍 · 정준영 · 이인범

포항공과대학 화학공학과 공정산업의 지능자동화 연구센터/  
산업과학기술연구소 이공부문 화공연구부  
(1993년 5월 26일 접수, 1993년 8월 2일 채택)

## Modification of the Joback Group Contribution Method Using Artificial Neural Networks

Kun-Hong Lee, June Young Jung and In Beum Lee

Automation Research Center, Department of Chemical Engineering,  
Pohang Institute of Science and Technology/  
Chemical Engineering Division, Research Institute of Industrial Science and Technology  
(Received 26 May 1993; accepted 2 August 1993)

### 요 약

인공지능의 기법을 활용하여 Group Contribution Method의 정확도를 향상시키는 일반적인 방법을 제안하였다. Joback GCM을 선택하고, 순수한 물질 400여개의 데이터를 사용하여 역전파 신경회로망을 30000-40000회 학습시켰다. 학습된 가중치 매트릭스를 사용하여 임계성질, 끓는 점, 어는 점, 생성엔탈피, 생성 Gibbs 자유에너지 및 이상기체의 비열을 예측해 본 결과, 기존의 수학적 모델을 이용하여 개발된 GCM보다 정확한 예측을 보임을 확인하였다.

**Abstract**—Using the technique of artificial intelligence, the general method to improve the accuracy of Group Contribution Methods was proposed. The Joback GCM was chosen and the back propagation neural networks were trained 30000 to 40000 times with the experimental data of more than 400 pure substances. The estimations of the critical properties, the normal boiling and the freezing temperatures, heats of formation, Gibbs free energies of formation, and ideal gas state heat capacities using trained weight matrices were much better than those from the conventional GCM developed on the basis of an assumed mathematical model.

### 1. 서 론

화학제품의 가격은 많은 인자에 의하여 결정되지만, 분리 및 정제의 비용이 가장 큰 부분을 차지하며[1], 그 중 분리공정의 설계는 대상이 되는 혼합물의 물성 데이터를 기초로 하여 이루어진다. 물성데이터는 실험적으로 얻어지기도 하나, 때로는 예측된 물성치를 사

용하기도 한다. 특히, 신물질의 합성하거나 다성분 혼합물을 다루게 되는 경우에는, 실험치가 존재하지 않는 경우가 보통이므로 물성치를 예측하는 신뢰성있는 방법이 요청된다.

물성치를 예측하는 방법으로 다양한 형태의 상관관계식들(correlations)이 제안되어 있으나, 가장 적용범위가 넓은 방법으로는 Group Contribution Method(이

하 GCM)를 들 수 있다. GCM은 주어진 작용기(functional groups)를 조합하여 만들 수 있는 모든 분자의 물성치를 추정할 수 있으므로, 제한된 종류의 분자들에만 적용되는 여타의 관계식들 보다 활용범위가 매우 넓다. GCM의 이러한 특성으로 인하여, 분자설계에 의한 신물질의 물성치 예측은 주로 GCM에 의존하고 있다[2-5]. 그러나, GCM은 일반적으로 정확도가 떨어지는 단점이 있으며, 예측치의 정확도가 물질에 따라서 불규칙하게 변화하므로 경우에 따라서는 매우 부정확한 예측치를 제공한다[6].

본 연구에서는 GCM의 정확도를 증가시킬 수 있는 새로운 방법을 제시하였으며, 최근 급속한 발전을 보이고 있는 인공지능(artificial intelligence)의 기법을 도입하여 Joback GCM의 정확도를 크게 향상시킬 수 있었다.

## 2. 본 론

GCM에서는 분자를 표현하는 작용기의 집합을 찾고, 각 작용기의 특성치를 합하여 분자의 특성치를 구한 후, 별도의 상관관계식에 대입하여 분자의 물성치들을 얻게 되므로 물성치의 예측에 영향을 주는 요소는 3가지가 있다. 첫째는, 어떠한 작용기를 선택하느냐는 것이다. 이는 각 GCM마다 나름대로의 근거를 가지고 있으며, 어느 것이 더 합리적인지를 판단하기는 어려우나, 주어진 작용기들로부터 다양한 분자들을 쉽게 조립할 수 있어야 한다. 둘째는, 선택된 작용기들의 특성치들이다. 지금까지 제안된 GCM들에서는 각 작용기의 특성치는 다수의 실험데이터를 사용한 다변수 선형 회귀분석(multivariate linear regression)을 통하여 결정되었다. 셋째는, 분자의 특성치와 분자의 물성치간의 상관관계식이다. 수학적식을 이용하여 대상계(system)를 모델링할 때는 그 계의 변수와 특성간의 상관관계의 기본 함수 형태를 미리 가정한 후에 최적의 매개변수를 결정하는 방법을 이용하고 있다. 그러나, 항상 타당한 가정을 통해서 기본식을 세우는 일은 사실상 불가능하기 때문에 이 가정으로부터 벗어나는 분자들에 대해서는 매우 나쁜 결과를 준다. 이 상관관계식에는 보다 정확한 예측을 위하여 작용기 집합에 대한 정보 외에 분자량이나 다른 물성치를 입력변수로 취하는 경우가 대부분이며 또 이웃하는 그룹의 종류에 따라서 다른 특성치를 갖게 하는 방법들이 적용되었다[7, 8].

GCM의 정확도에 영향을 주는 3가지 요소는 서로 강하게 연관되어 있으므로, 이들 요소들을 통합적으로 다룰 수도 있다. 즉, 각각의 작용기들을 변수로 보고 특정한 상관관계를 가정하지 않은 상태에서 실험으로 얻어진

물성치 데이터를 사용한 최적화를 통하여 각 작용기의 특성치들을 구하는 것이다. 이는 인공지능의 한 분야로서 최근 관심을 끌고 있는 신경회로망(neural network)을 사용함으로써 달성할 수 있다.

역전파(back propagation) 신경회로망은 매우 복잡한 대상계의 변수와 그 계의 특성을 관계짓는데 유용함이 많은 적용 사례를 통하여 증명된 바 있다. 역전파 신경회로망이 가정을 통해서 세워진 제한된 형태의 수학적 모델식보다 좋은 성능을 보이는 것은 신경회로망 모델이 임의의 상관관계를 나타낼 수 있다는데 기인한다. 두 개의 성분으로 구성된 입력 벡터를 갖는 신경회로망 모델이 입력 벡터와 출력층의 출력을 축으로 하는 3차원 좌표상에서 보이는 표면(surface)은 연결강도에 따라서 임의적으로 변화한다[9]. 또, 세 개의 층(input layer, hidden layers, output layer)으로 이루어진 역전파 신경회로망을 통하여 임의의 정성적인 관계를 표현할 수 있고, 네 개의 층(input layer, 2 hidden layers, output layer)으로 이루어진 역전파 신경회로망을 통하여 임의의 정량적인 상관관계를 표현할 수 있음이 수학적으로 증명되었다[10].

본 연구에서는 지금까지 제안된 GCM중 그 활용범위가 가장 넓은 Joback GCM을 대상으로, 그 작용기에 부여된 특성치를 신경회로망(neural network)을 사용하여 최적화함으로써 물성치 예측의 정확도를 향상시켰다.

### 2-1. Joback GCM

Joback은 Lydersen의 GCM을 수정, 확장하여 새로운 GCM을 제시하였으며[8], 이것은 Ambrose의 GCM에 비하여 훨씬 간단하면서도 비슷한 정도의 정확성을 보였다[11]. Joback GCM은 작용기를 다른 GCM에 비하여 간단하게 나누었으면서도 많은 수의 물질들을 표현할 수 있는 장점이 있다. Joback의 GCM은 40개의 작용기를 가지고 있으며, 이들 작용기에는 각각 특성치  $\Delta$ 가 주어져 있다. 주어진 분자의 물성치를 계산하기 위해서는, 먼저 분자를 작용기의 조합으로 분해한 후 각 작용기의 특성치들로부터 분자의 특성치  $\Sigma\Delta$ 를 얻게 된다. 이렇게 얻어진 분자의 특성치를 다음과 같은 상관관계식을 대입하여 물성치를 얻는다.

$$\begin{aligned} P_r &= [0.113 + 0.0032 n_w - \Sigma\Delta P]^{-2} \\ T_r &= T_b [0.584 + 0.965 \Sigma\Delta T + (\Sigma\Delta T)^2]^{-1} \\ V_r &= 17.5 + \Sigma\Delta V \\ T_b &= 198 + \Sigma\Delta b \\ T_f &= 122 + \Sigma\Delta f \\ \Sigma H_f &= 68.29 + \Sigma\Delta H \\ \Sigma G_f &= 53.88 + \Sigma\Delta G \end{aligned}$$

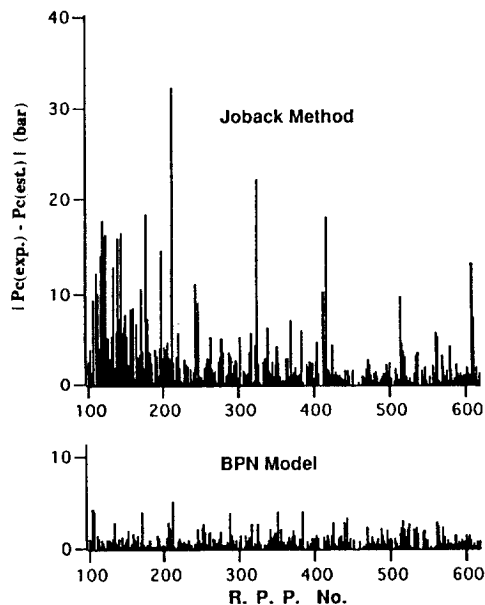


Fig. 1. Absolute errors in critical pressure estimation.

$$C_p^* = (\Sigma \Delta a - 37.93) + (\Sigma \Delta b + 0.210)T \\ + (\Sigma \Delta c - 0.000391)T^2 + (\Sigma \Delta d + 0.000000206)T^3$$

한편, Joback GCM은 몇 가지 단점을 갖고 있다. 첫째는, 다양한 분자들에 대하여 고른 정확성을 보장하지 못한다는 점이다. Fig. 1은 425개의 물질에 대하여 GCM에 의한 임계 압력 예측에서의 각 물질의 절대 오차를 나타낸다. 여기서 R. P. P.'s No.는 참고문헌[11]의 물질치 데이터 베이스에 있는 각 화합물의 번호이며, BPN은 역전파 신경회로망을 사용한 경우를 의미한다. Joback GCM에 의한 임계압력 예측의 평균 절대 오차는 2.41(bar)이고, 평균 상대 오차는 4.13%로 425개의 물질에 대한 평균적인 예측 성능은 우수하다고 할 수 있다. 그러나 각각의 물질에 대한 예측치와 실험치의 오차를 고려하여 보면 대부분의 물질은 5 bar이하의 타당한 예측치를 제공하지만 나머지 물질들에 대해서는 지나치게 큰 오차를 보이기 때문에 예측치로서의 가치가 없는 경우도 많이 발생함을 볼 수 있다. 이것은 모든 GCM의 공통된 약점이며, 따라서 Joback법을 안심하고 사용하기 위해서는 모든 물질에 대해서 고르게 정확도를 확보하도록 수정이 가져져야 할 것이다.

둘째로, Joback GCM은 임계온도나 임계압력의 계산에서 끓는 점( $T_b$ )이나 원자의 갯수( $n_a$ ) 등, 분자구조에 대한 정보 외에 부가적인 정보를 필요로 한다. Fig. 2는 임계온도의 예측시에 입력 변수로 사용되는  $T_b$ 에, 실험

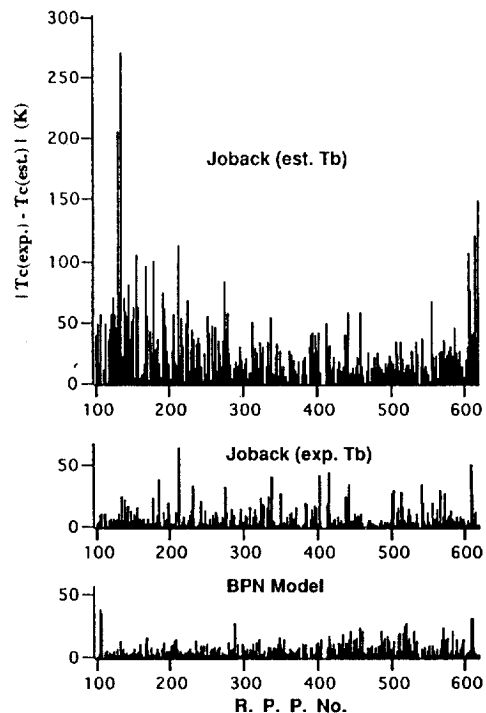


Fig. 2. Absolute errors in critical temperature estimation.

데이터를 사용한 경우와 Joback 방법에 의한 예측치를 이용한 경우에 대한 비교를 나타내었다. Joback방법의 끓는 점 예측식은 매우 간단하지만 그 예측치는 평균 절대 오차가 14.37 K, 평균 상대 오차가 2.99%로 다른 물질치 예측 오차에 비해 지나치게 큰 오차를 갖고 있다. 이 예측된 끓는 점을 이용한 임계 온도 예측은 평균 절대 오차가 21.6 K, 평균 상대 오차가 4.08%로 끓는 점 예측보다 더 큰 오차를 보인다. 그러나 실험으로부터 얻은 끓는 점 데이터를 이용한 경우에는 평균 절대 오차가 5.91 K, 평균 상대 오차가 1.01%의 오차를 보임으로써 훨씬 정확한 예측치를 제공한다. 여기서 알 수 있는 것은 Joback방법에서 임계온도 예측의 정확도는 그룹에 대한 정보 외에 끓는 점의 정확도에 상당히 큰 영향을 받는다는 사실이다. 즉, 1차 예측치( $T_b$ )에 포함된 오차는 2차 예측치( $T_c$ )에서 상쇄 또는 누적되므로, 물질에 따라서는 매우 큰 오차를 나타낸다.

## 2-2. 역전파 신경회로망

역전파 신경회로망은 다수의 소자(node)를 가진 입력층(input layer), 중간층(hidden layer) 및 출력층(output layer)으로 이루어진다. 같은 층의 처리 소자들은 같은 전달함수를 가지며, 동일한 학습방법을

**Table 1. Functional groups used in the Joback GCM**

No. Group	No. Group	No. Group
Nonring	Halogen	Nitrogen
1 -CH <sub>3</sub>	16 -F	30 -NH <sub>2</sub>
2 >CH <sub>2</sub>	17 -Cl	31 >NH(nonring)
3 >CH-	18 -Br	32 >NH(ring)
4 >C<	19 -I	33 >N-(nonring)
5 =CH <sub>2</sub>	Oxygen	34 -N=(nonring)
6 =CH-	20 -OH(alcoh.)	35 -N=(ring)
7 =C<	21 -OH(phenol)	36 -CN
8 =C=	22 -O-(nonring)	37 -NO <sub>2</sub>
9 ~CH	23 -O-(ring)	surfur
10 ~C-	24 >C=O(nonring)	38 -SH
Ring	25 >C=O(ring)	39 -S-(nonring)
11 -CH <sub>2</sub> -	26 O=CH-(aldehyde)	40 -S-(ring)
12 >CH-	27 -COOH(acid)	
13 >C<	28 -COO-(ester)	
14 =C-	29 =O(except above)	
15 =C<		

note: ~ denotes triple bond.

통하여 연결강도가 조절된다.

역전과 신경회로망의 중요성은 하위층의 각 처리 소자의 출력형태에 따라서 상위층의 특정처리 소자가 활성화된 상태를 띠는 점이다. 즉, 하위층의 출력 패턴이 갖는 여러 가지의 특성은 상위층의 활성화된 처리 소자와 대응을 이룬다. 이러한 입력 패턴의 특성은 사람이 구분할 수 있는 없든 간에 중간층의 각 처리 소자의 출력에 의해서 추출된다[12]. 이 특성 추출(feature extraction) 기능은 각 본자를 이루는 그룹들에 대한 정보를 입력으로 하고, 그 본자의 물성치를 출력으로 하여 신경회로망을 학습시켰을 경우에 본자의 구성 작용기의 패턴에 따른 물성치의 특성을 스스로 학습하도록 함으로써 입력의 수치적인 처리 뿐만 아니라 그 패턴에 대한 분석도 그 물성치의 예측치를 결정하는 정보로 사용되도록 하여 보다 정확한 예측을 가능하도록 한다.

역전과 신경회로망의 학습은 입력된 예제 패턴에 대해 그와 대응되는 출력을 나타내는 최적의 연결 강도를 찾는 과정이다. 역전과 신경회로망의 학습

알고리즘들은 최소 자승법이 반복 계산형태로 변형된 것이라고 할 수 있다. 이 변환된 알고리즘들은 최대 경사법(steepest descent method) 또는 델타법(delta rule)이라고 한다. 본 연구에서는 델타법을 여러 층의 신경회로망에 대하여 확장한 일반화된 델타법(generalized delta rule)[12]을 사용하였다.

### 2-3. 적용된 신경회로망 모델

각각의 물성치 예측을 위하여 사용된 역전과 신경회로망의 구조는 다음과 같다. 입력층의 처리 소자는 각 물성치마다 43개로 구성되어 있다. 분자 구성 그룹의 정보는 Joback GCM에서 제안된 40개의 그룹을 기본으로 하였다. 그리고 cis-, trans-이성질체와 ortho-, meta-, para-이성질체를 구분하는 2개의 처리 소자를 추가하였다. 여기에 항상 1을 갖는 잡음신호(bias)항을 합해서 모두 43개의 입력층 처리 소자를 구성하였다. 그룹에 대한 정보를 받는 40개의 처리 소자는 각각 특정그룹의 갯수를 입력받게 하였다. Joback방법에서 제안한 40개의 그룹은 Table 1에 나타난 번호대로 입력처리 소자에 그 갯수가 입력된다. Table 2는 학습예제에서 몇몇 물질에 대한 입력화일을 나타낸다. 첫 번째 숫자는 참고문헌[11]의 물성치 자료 베이스에서의 각 물질에 대한 번호를 나타낸다. 2-41열까지의 숫자는 각 물질을 표현하는데 필요한 각각의 그룹의 갯수를 나타낸다. 첫 번째 물질은 chlorodifluoromethane( $\text{CHClF}_2$ )이므로 3번 그룹하나, 16번 그룹 둘, 17번 그룹 하나로 표시됨을 나타낸다. 42 및 43번째 숫자는 cis-, trans-이성질체와 ortho-, meta-, para-이성질체의 구분을 표시한다. 418번과 419번은 40개 작용기에 대한 입력은 동일하나 42번째 열이 cis형은 0, trans형은 1로 입력된다. 607번부터 609번은 43번째 열이 서로 다른 값을 가지는데, 여기서 0, 1, 2는 각기 ortho-, meta-, para-이성질체를 나타낸다. 이성질체가 없는 물질들은 이 두 처리 소자의 입력으로 0을 취하도록 학습시켰다. 제일 마지막 열은 잡음신호항이며 항상 1을 입력한다.

중간층은 각 물성치에 대한 신경회로망마다 서로 다른 갯수의 처리 소자로 구성되어 있는데, 이로부터

**Table 2. Example of inputs to input layer**[illegible]

Table 3. The structure of artificial neural network models

Number of nodes	T <sub>c</sub>	P <sub>c</sub>	V <sub>c</sub>	T <sub>b</sub>	T <sub>f</sub>	ΔH <sub>f</sub>	ΔG <sub>f</sub>	C <sub>p</sub> *
Input layer	43	43	43	43	43	43	43	43
Hidden layer 1	20	12	14	14	19	23	25	35
Hidden layer 2	0	0	0	0	0	18	12	23
Output layer	1	1	1	1	1	1	1	1
Number of data used for training	465	425	279	467	343	337	287	351

기존의 GCM에서도 각 물성치와 그룹이 갖는 상관관계식이 서로 다른 기본식을 바탕으로 세워졌듯이 신경회로망에서도 다른 형태의 상관관계 표면을 이룬다는 것을 알 수 있다. 중간층의 처리 소자의 개수는 학습시키는 과정에서 시행착오를 거쳐서 수렴성이 좋은 방향으로 정해졌다. 신경회로망을 통해서 정성적 처리를 하는 경우에는 이진수로 입력과 출력을 나타내는데, 알려진 경험법칙에 의하면 이 때 중간층의 처리 소자의 개수는 일반적으로 입력층의 처리 소자의 개수의 1/4정도로 구성하도록 되어 있다[12]. 그러나 본 연구에서는 정량적인 처리에 이용되었으므로 이 들보다는 많은 수의 처리 소자가 필요하였다. 우선은 약 12개의 처리 소자로 학습을 시작하여 학습이 쉽게 되는 경우에는 처리 소자를 감소시키고, 수렴이 되지 않은 상황에서 학습이 더 이상 진전되지 않는 경우에는 처리 소자의 개수를 증가시키는 과정을 통하여 각 신경회로망 모델의 중간층 처리 소자의 개수가 결정되었으며, Table 3에 결과를 나타내었다.

모든 모델에서 출력층은 한 개의 처리 소자를 가진다. 본 연구에서 이용된 모든 신경회로망의 처리 소자는 전달 함수로서 다음 식을 이용하였다.

$$y = \frac{1}{1 + \exp[-x]}$$

이 식은 각 처리 소자가 항상 0과 1 사이의 값을 출력하도록 한다. 그러므로 학습시에 기대값으로는 모든 학습 예제에서 최대값과 최소값이 각각 0과 1에 접근하도록 하고, 나머지는 이들 사이에 분포하도록 하는 표준화된 값을 이용하였다. 출력 처리 소자에서 출력된 값들은 다시 표준화 과정의 역과정을 거쳐서 각 물성치를 제공하도록 되어 있다. 이 방법에서는 어떤 분자의 표준화된 물성치가 1보다 큰 경우는 있을 수 없다. 따라서 학습에 이용된 예제의 최대값이나 최소값에 비해서 지나치게 크거나 작은 물성치를 갖는 분자에 대해서는 좋은 결과를 보일 수 없겠으나 실제 학습에서 충분히 많은 예제를 이용하였기 때문

Table 4. Comparison of accuracies in the estimation of physical and thermodynamic properties of pure substances

	Joback method		BPN model	
	AAE	AAPE	AAE	AAPE
P <sub>c</sub> (bar)	2.41	4.13	0.72	2.18
T <sub>c</sub> (K; est.T <sub>b</sub> )	21.26	4.08	5.50	0.99
T <sub>c</sub> (K; exp.T <sub>b</sub> )	5.91	1.01		
V <sub>c</sub> (cm <sup>3</sup> /mole)	9.85	6.16	4.62	1.56
T <sub>b</sub> (K)	14.37	2.99	3.92	1.05
T <sub>f</sub> (K)	22.07	11.07	0.88	0.45
H <sub>f</sub> (KJ/mole)	10.48	16.78	2.03	5.29
G <sub>f</sub> (KJ/mole)	7.91	14.39	1.50	6.21
C <sub>p</sub> *(298 K)	5.52	4.12	5.15	4.32
C <sub>p</sub> *(373 K)	7.14	4.31	6.28	4.44
C <sub>p</sub> *(500 K)	13.27	6.42	8.24	4.73
C <sub>p</sub> *(800 K)	48.07	17.70	12.89	5.66

AAE: Average Absolute Error

AAPE: Average Absolute Percent Error

C<sub>p</sub>\*: constant pressure heat cappacity at ideal gas state, J/mol K

에 표준화된 출력의 범위를 벗어나는 경우는 많지 않을 것이다.

각 물성치를 위한 신경회로망 모델의 학습은 참고 문헌[11]의 물성치 자료 베이스에서 실험치를 얻을 수 있는 물질들을 예제로 사용하였다. 각 물성치를 학습시키는데 사용된 실험데이터의 개수도 Table 3에 나타내었다. 각 물성치의 예측 모델을 학습시키는데는 모델마다 차이는 있으나 약 30000 내지 40000회의 학습이 반복되었다. 학습이 끝난 신경회로망은 가중치 매트릭스(weight matrix) W의 형태로 주어지며, 이 가중치 매트릭스로부터 다음과 같은 식을 통하여 예측치를 얻게 된다.

$$Q = Fo[\sum_k W_{ok} F2_k \{ \sum_j W_{kj} F1_j (\sum_i W_{ji} I_i) \}]$$

여기서 Q는 각 물성치의 정규화된(normalized) 출력 값(즉, T<sub>c</sub>, P<sub>c</sub>, V<sub>c</sub> 등)을 나타내며, Fo, F1 및 F2는 각각 출력층, 첫번째 중간층 및 2번째 중간층의 각 node에서 사용되는 전달함수이다. I는 Joback Table의 각 작용기의 개수이며, W<sub>ok</sub>, W<sub>kj</sub> 및 W<sub>ji</sub>들은 각 node간의 연결 강도를 나타낸다.

## 2-4. 결과 및 검토

Table 4는 물성치 예측을 위해 새로 제안된 신경회로망 모델과 기존의 Joback GCM의 성능을 비교하기 위하여, 실험치가 있는 물질들에 대해서 두 모델에 의한 각 물성치의 예측결과와 실험치와의 오차를 비

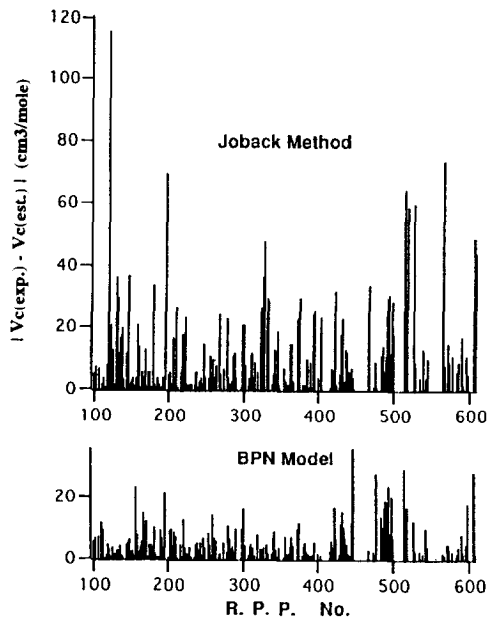


Fig. 3. Absolute errors in critical volume estimation.

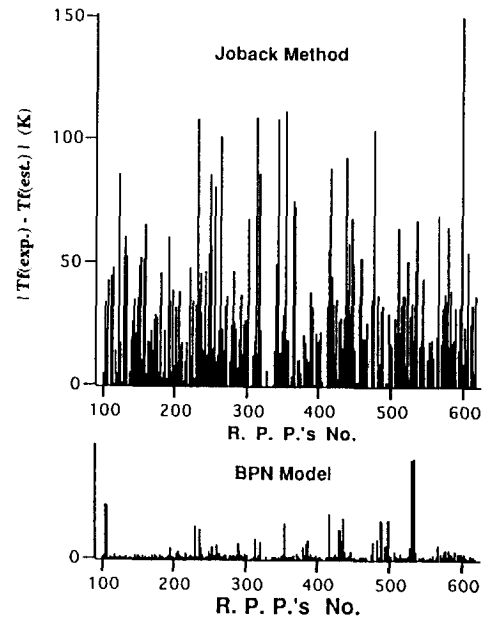


Fig. 5. Absolute errors in freezing temperature estimation.

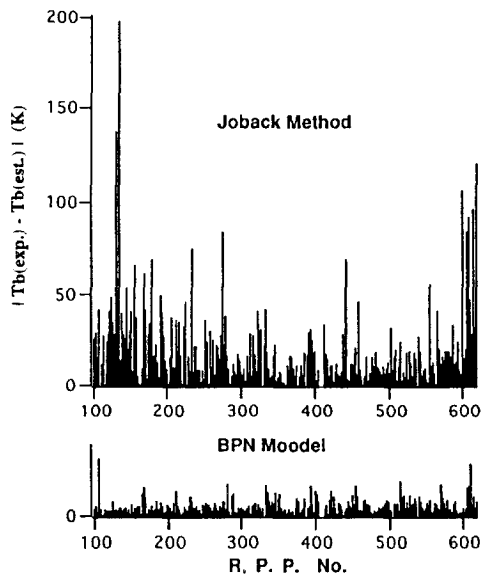


Fig. 4. Absolute errors in normal boiling temperature estimation.

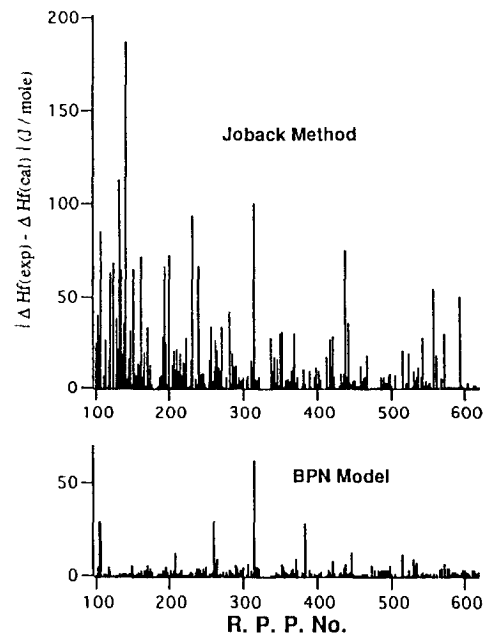


Fig. 6. Absolute errors in the estimation of standard enthalpy of formation for ideal gas at 298.2 K.

교한 것이다. 본 연구에서 고려한 모든 물성치들의 예측에서 신경회로망 모델이 더 정확함을 알 수 있다.

예측된 물성치를 사용하는 유저의 입장에서는 많은 물질에 대한 평균적인 오차보다는 특정한 물질에 대한 오차가 더 중요하다. Fig. 1-8은 GCM이 물질에 따라서

예측치의 정확도가 크게 차이가 남을 분명히 보여준다. 또한, 새로운 신경회로망 모델과 Joback방법의 예측 결과의 실험치와의 절대 오차를 막대 그래프 형태로

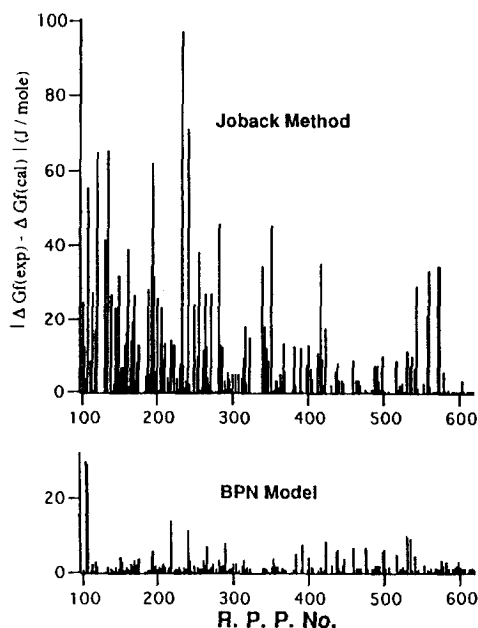


Fig. 7. Absolute errors in the estimation of standard Gibbs free energy of formation for ideal gas at 298.2 K and 1 atm.

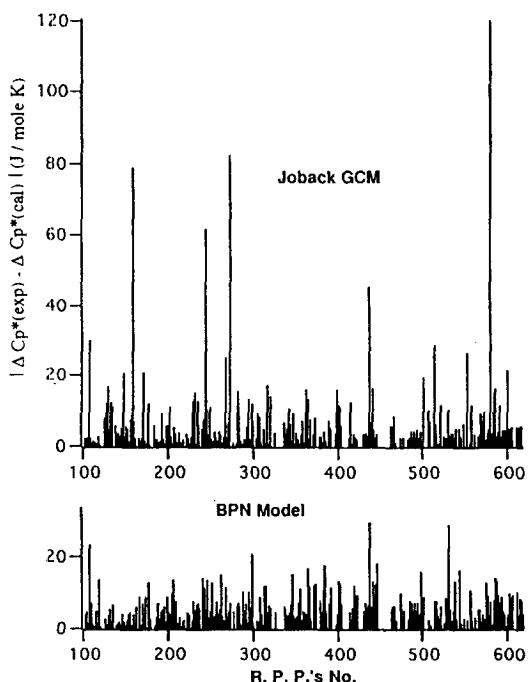


Fig. 8. Absolute errors in the estimation of isobaric heat capacity of ideal gas at 298.2 K.

Table 5. Estimation of freezing temperatures of isomers

Experimental data(K)	Joback method		BPN model		
	Estima- tion(K)	AAE (K)	Estima- tion(K)	AAE (K)	
1,2-dimethylcyclo- pentane(C <sub>7</sub> H <sub>14</sub> )					
cis-	219.3	174.8	44.5	219.7	0.4
trans-	155.6	174.8	19.2	155.3	0.3
terphenyl(C <sub>18</sub> H <sub>14</sub> )					
ortho	330.0	383.9	53.9	329.8	0.2
meta	360.0	383.9	23.9	359.8	0.2
para	485.0	383.9	101.1	367.9	117.2

AAE: Absolute Average Error

비교함으로써 두 방법의 성능을 한눈에 알아볼 수 있도록 하였다. Fig. 1-8에서 보는 바와 같이 새로 제안된 신경회로망 모델은 기존의 다변수 선형 회귀분석법으로 고안된 Joback GCM보다 월등한 성능을 보임을 알 수 있다.

Joback GCM에서는 그룹에 대한 정보 외에 임계온도를 계산하는데 끓는 점이 변수로 필요하고, 임계압력을 계산하는데 원자의 갯수를 필요로 하는 등 정확한 예측을 위해서 보다 복잡한 식을 이용하였으며 Fig. 2에서 나타난 바와 같이 임계 온도의 정확성은 끓는 점의 정확성에 상당히 크게 의지하고 있다. 그러나 새로이 제안된 신경회로망 모델에서는 그룹에 대한 정보만으로도 훨씬 정확한 예측을 가능케 함을 확인하였다.

Joback GCM에서는 이성질체들이 구별되지 않지만, 신경회로망 모델에서는 이를 구별하기 위한 소자를 첨가하였으므로 각 이성질체별로 다른 물성치를 예측할 수 있다. 어느 점에서는 분자간의 간격이 조밀하여 다른 물성치들에 비하여 이성질체의 영향이 크게 나타나므로 이성질체를 다룰 때 기준으로 삼는 것이 타당하다. Table 5는 cis/trans 및 ortho/meta/para 이성질체들의 어느 점 예측의 예이며, 이 경우 신경회로망은 이성질체의 특성을 반영할 수 있음을 보여준다.

신경회로망의 정확한 예측치는 상수의 증가에 의한 것이다. 예를 들면, Joback GCM으로 임계압력을 예측하기 위해서는 40개의 상수가 필요하나, 신경회로망의 경우는 860개 (42×20+20×1)의 상수가 필요하다. 그러나, 이들은 상수(constants)이므로 일단 결정이 되면 계산시간이 약간 더 필요하다는 점외에는 다른 어려움을 수반하지 않는다. 현재 상용화되어 있는 개인용 컴퓨터(IBM 386)를 사용하여 계산해 본 결과, 시간 차이는 거의 없었다.

### 3. 결 론

GCM은 물성치 예측법으로 가장 이상적이거나, 예측치의 정확도가 떨어짐으로 인하여 널리 활용되지 못하고 있다. 본 연구에서는 모든 GCM의 정확도를 크게 증가시킬 수 있는 일반적인 방법과 그 예를 제시하였다. Joback GCM을 대상으로 선택하고 신경회로망을 사용하여 작용기 특성치의 최적화를 수행하였으며, 그 결과로 기존의 GCM보다 훨씬 정확하게 임계물성치를 예측할 수 있었다. 학습이 끝난 신경회로망은 숫자들로 구성된 단순한 매트릭스이므로, 개인용 컴퓨터만 있으면 누구나 쉽게 사용할 수 있다. 본 연구에서는 소규모의 데이터 베이스를 활용하여 계산 예를 보였으나, 이 방법을 DIPPR이나 Dortmund Data Bank 등의 대규모 데이터 베이스와 연계할 경우 상당히 정확한 GCM을 개발할 수 있을 것으로 예상된다.

다수의 변수를 변화시키면서 최적화점을 찾는 일반적인 알고리즘은 확립되어 있지 않다. 역전파 신경회로망은 이러한 경우에 적용할 수 있는 방법중에 하나이며, GCM과 같이 다수의 상수를 결정하여야 할 경우에는 매우 적절한 방법이라고 판단된다.

### 감 사

본 연구를 지원해 주신 산업과학기술연구소 및 공정산업의 지능자동화연구센터에 감사드립니다.

### 사용기호

$C_p^*$	: isobaric heat capacity of the ideal gas
$F_o$	: transfer function in the output layer
$F_1$	: transfer function in the 1st hidden layer
$F_2$	: transfer function in the 2nd hidden layer
$I$	: number of $i$ th functional groups in a molecule
$P_c$	: critical pressure
$Q$	: arbitrary physical property
$T_b$	: normal boiling temperature
$T_c$	: critical temperature
$T_f$	: freezing temperature
$V_c$	: critical volume
$W$	: weight matrix between layers
$X$	: input of transfer function
$Y$	: output of transfer function
$\Delta H_f^\circ$	: standard enthalpy of formation for the ideal gas at 298.2 K
$\Delta G_f^\circ$	: standard Gibbs free energy of formation for

the ideal gas at 298.2 K and 1 atm

### 그리스 문자

$\Delta$	: difference or characteristic value of a functional group
$\Sigma$	: summation

### 상첨자

$o$	: properties at standard state
$*$	: properties at ideal gas state

### 하첨자

$b$	: properties at boiling point
$c$	: properties at critical point
$f$	: formation or properties at freezing point
$i$	: summation index
$j$	: summation index
$k$	: summation index
$p$	: constant pressure

### 참고문헌

1. Sherwood, T. K., Pigford, R. L. and Wilke, C. R.: "Mass Transfer", McGraw-Hill, New York, NY (1975).
2. Gani, R. and Brignole, E. A.: *Fluid Phase Equilibria*, **13**, 331(1983).
3. Joback, K. G. and Stephanopoulos, G. in Sirola, J. J., Grossmann, I. E. and Stephanopoulos, G. ed.: "Foundations of Computer-Aided Process Design", Elsevier, Amsterdam(1990).
4. Nielsen, B., Gani, R. and Fredenslund, A. in Bussemaker, H. Th. and Ledema, P. D. ed.: "Computer Applications in Chemical Engineering", Elsevier, Amsterdam(1990).
5. Gani, R., Nielsen, B. and Fredenslund, A.: *AIChE J.*, **37**(9), 1318(1991).
6. Jung, J.-Y.: MS Thesis, POSTECH, Pohang, Korea (1993).
7. Klinecicz, K. M. and Reid, R. C.: *AIChE J.*, **30**(1), 137(1984).
8. Joback, K. G. and Reid, R. C.: *Chem. Eng. Comm.*, **57**, 233(1987).
9. Lapedes, A. and Farber, R.: "How Neural Nets Works", Neural Information Processing Systems, Denver, CO, 1987, American Institute of Physics (1988).
10. Bhat, N. and McAvoy, T. J.: *Computers Chem. Eng.*,



- 14(4/5), 573(1990).
11. Reid, R. C., Prausnitz, J. M. and Poling, B. E.: "The Properties of Gases and Liquids", 4th ed., McGraw-Hill, New York, NY(1987).
  12. Freeman, J. A. and Skapura, D. M.: "Neural Network Algorithms, Applications and Programming Techniques", Addison-Wesley, Reading, MA(1991).