

# 신경망을 기반으로 한 새로운 물성 예측을 통한 가스공정의 위험성 평가

유미정, 김용하, 김구회\*, 서정철\*\*, 윤인섭  
서울대학교 응용화학부, 서울대학교 화학공정신기술연구소\*  
한국가스공사 연구개발원\*\*

## Thermal hazard assessment from the estimation of physical properties based on Neural networks

Mi Jung Ryu, Young Ha Kim, Jung Chul Suh\*, Ku Hwoi Kim\*\*, En Sup Yoon  
School of chemical Engineering, Seoul National Univ. Institute of Chemical Process\*  
R&D Division, Korea Gas Corporation\*\*

### 1. 서론

새로운 화학 공정을 설계할 때, 공정에 대한 위험성 평가는 필수적이다. 특히, 예기치 않은 부가 반응으로 인한 급격한 온도의 상승은 runaway reaction과 같은 폭발적인 반응을 일으켜 물질적인 피해뿐만 아니라 사상자를 발생시키기도 한다. 따라서 이러한 높은 반응성으로 인한 위험 요인은 공정 설계시 반드시 고려되어야 한다. 위험성 평가의 선별 방법으로서 반응에 의한 방출열량의 예측은 좋은 기준이 되어 왔다. 전통적으로 반응열은 반응에 포함된 물질의 물성을 통해 계산 될 수 있으나, 새로운 물질이나 고분자 물질이 포함된 반응에 대해서는, 물성치가 알려지지 않은 경우가 많기 때문에, 대개 직접적인 실험을 통해서만 측정 되어 왔다. 그러나 실험적 접근 방식은 ARC, DSC 등의 값 비싼 분석 장치를 요구할 뿐만 아니라 실험 결과로부터의 분석의 어려움, scale-up error, 그리고 실험 자체가 가진 잠재적 위험성 등의 문제를 안고 있다. 따라서 적절한 이론적 접근 방법은 경제적이고 안전한 공정 설계에 중요한 역할을 할 수 있다.

새로운 이론적 방식으로 보다 정확하고, 다양한 물질에 적용할 수 있는 향상된 모델을 제시했다. 제안된 모델은 기본적으로 Molecular structure에 기반 한 group contribution method이며, 다층신경망(Multiple Layer Perceptron)을 통해

학습시키고, 여러 가지 통계이론을 적용시켜 예측 성능을 향상 시켰다.

예측된 물성 값은 공정에 포함된 반응열 계산에 적용되어 정량적 위험성 평가에 활용될 수 있다.

## 2. 이론

### 물성 예측 방법

가장 대표적인 방법인 Group Contribution Method는 물질의 물성치가 그 물질을 구성하는 기능기(Functional group)에 따라 결정된다는 가정을 바탕으로 하는 방법으로 계산이 상대적으로 쉽고 비교적 좋은 예측 성능을 보인다. Joback은 물질의 40개의 기능기로 분석하여 회귀 모델을 제안하였고, Constantinou는 Joback이 제안한 방법에 second order functional group을 추가하였다. Lee는 Joback이 정의한 group 정보를 사용하여, 신경회로망 기법을 이용한 방법을 제시했으나, 데이터베이스를 모델의 학습에만 사용하고, 평가나 테스트 과정을 거치지 않았으므로, 주어진 학습데이터에만 과적합(Over-fitting)된 결과로서 일반화 과정을 거쳐야 한다고 보인다.

그러나 기존의 group contribution methods는 물질에 따라 예측의 정확성에 큰 차이를 보였다. 특히 peroxide를 포함한 물질이나, ring 구조를 가진 물질일 경우에 정확성이 크게 떨어지는 현상을 보였다. 이는 초기에 Joback이 사용한 Raw dataset이 400여개로 제한되어 있고, Raw data에 포함된 물질이 또한 그가 사용한 40개의 기능기를 대표하지 못했기 때문으로 보인다.

다른 방법으로 통계열역학과 계산 양자 화학에 기반을 둔 예측 방법이 있다. GAUSSIAN이나 GAMESS와 같은 소프트웨어가 이미 상용화 되어 있고, group contribution method에 비해 상당히 정확한 예측이 가능하지만, 6-7개의 저분자 물질에만 적용 가능하고, 고분자 물질의 경우에는 계산이 불가능하다.

## 3. 모델링

### Preparing database

데이터로부터의 학습을 통한 모델링에서 가장 중요한 것은 데이터의 질 및 전처리이다. DIPPR database로부터 물성 값들을 알아내고, 이를 구조식으로 나타낸 후 각 물질의 기능기를 분석하여 데이터베이스를 확립했다. 총 물질의 수는 1400여개이며, 이는 이전에 Joback이 400여개의 물질만을 포함한 데이터에 기반한

것에 비해, 모델의 활용도와 신뢰도 향상에 도움이 된다고 할 수 있다.

입력 변수는 53개로 구성된 기능기(functional group)들과 1개의 분자량으로 모두 54개이며, 기능기에는 물질의 구조적 특성을 규명하는 정보가 포함된 기능기의 occurrence를 표현한 "interval" 변수와 기능기의 위치 정보를 고려한 "ordinary" 변수로 구성되어 있다. [Table 1] 출력 변수로는 The Ideal gas heat capacity, Enthalpy of formation at 298K, Gibbs Energy of formation at 298K 등 DIPPR 에 포함된 18개의 물성치이다.

기능기의 종류와 출현횟수, 위치정보로서 행렬형태의 입력 값들이 정해지고 각 기능기의 종류와 출현횟수에 따른 기여도의 비선형성을 고려해, Skewness 와 Kurtosis를 고려하여 데이터를 변환하고, 정규화 시킨다.

Table 1. 54개의 입력 변수

I - Ending Group			II - Middle Group			III - Ring Group	
-CH <sub>3</sub>	-SH	-COH	>C<	>N-	-SO-	CH <sub>2</sub>	N
=CH <sub>2</sub>	-Br	-COOH	>C=	=N-	o-B	CH	CO
≡CH	-F	=O	=C=	-NH-	m-B	-CH	O
≡N	-Cl	-OH	-C≡	-O-	p-B	>C	=C
-NH <sub>2</sub>	-I	=S	-CH <sub>2</sub> -	-S-	-CO <sub>2</sub> -	-C	S
-NO <sub>2</sub>	-phenyl	-H	>CH-	-CO-	* <sup>(1)</sup>	-N	R-C-R
IV - MW(분자량)			-CH=	-SO <sub>2</sub> -	** <sup>(2)</sup>	NH	R-CH-R

\*<sup>(1)</sup> : 3-branch를 가진 benzene의 구분(ordinary variable)

\*\*<sup>(2)</sup> : 4-branch를 가진 benzene의 구분(ordinary variable)

### **MLP Architecture**

MLP(Multi-Layer Perceptron)은 universal approximator로서, 입출력간의 관계를 모르는 경우에도 우수한 예측성능을 보인다. activation function으로는 'tanh' function(1), combination function으로는 Linear combination function(2)을 사용했다.

$$g(x) = \tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}} \quad (1)$$

(2)

MLP 에서 하나의 hidden unit 에 대한 activation formula는 다음 (3) 과 같다.

$$h_j = \exp \left[ -b_j^2 \sum_i (w_{ij} - x_i)^2 \right] \quad (3)$$

$h_j$ :  $M$ -estimation tuning constant  $b_j$ : the number of trials for  $(\text{ial})$  error

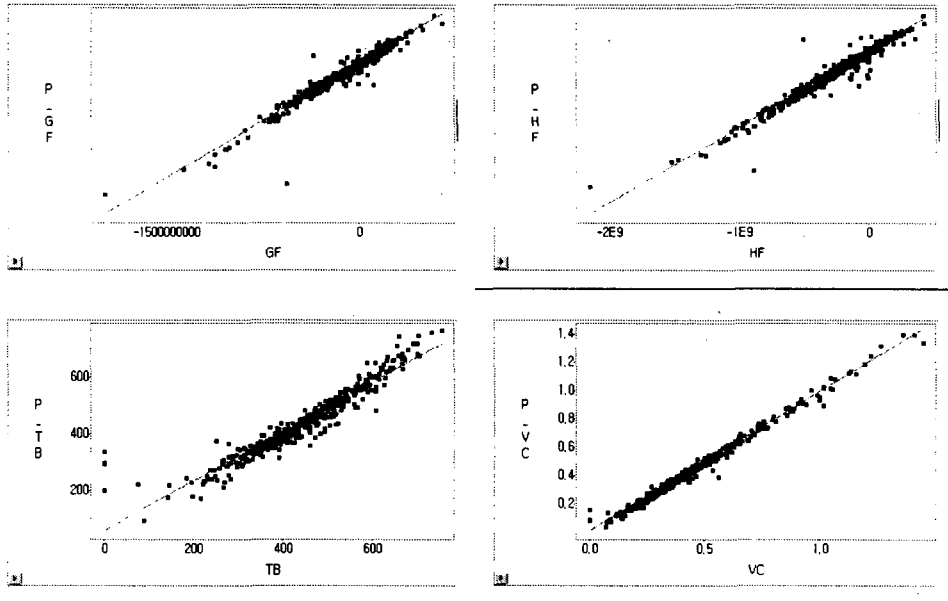
#### 4. 결과 및 토의

제안된 모델로 DIPPR 에 있는 데이터를 기반으로 이상 기체의 끓는점, 임계 압력, 임계온도 및 이상기체의 자유에너지, 형성 엔탈피, 엔트로피, 정압 열용량 등을 예측할 수 있다. [Table 2]

Table 2. 예측 가능한 물성

$T_B$	Boiling Point	$G_f$	Gibbs Energy of Formation
$P_C$	Critical Pressure	$H_C$	Enthalpy of Combustion
$T_C$	Critical Temp.	$H_f$	Enthalpy of Formation
$V_C$	Critical Volume	$H_{fu}$	Enthalpy of Fusion
$T_F$	Flash Point	S	Entropy
$T_{TP}$	Triple Point Temp.	$C_P$	Heat Capacity at Constant Pressure
$P_{TP}$	Triple Point Pressure	AF	Acentric Factor
$T_M$	Melting Point	$Z_C$	Critical Compressibility

제안된 모델은 Joback 과 Lee의 모델에 비해 정확한 결과를 보인다.



- \*  $G_f$  : Gibbs energy of formation at 298K(J/Kg/mol)
- \*  $H_f$  : Enthalpy of formation at 298K(J/Kg/mol)
- \*  $T_B$  : normal boiling point(K)
- \*  $V_C$  : critical volume (m3)

Figure 1. Fitness analyses in the estimation of Gf, Hf, Tb, Vc

Table 3. 다른 방법과 비교 분석

	Joback method	Lee's model	New model
Hf(J/Kg/mol)	16.78	5.29	0.91
Gf(J/Kg/mol)	14.39	6.21	0.97
Tb(K)	2.99	1.05	1.24
Tc(K)	4.08	0.99	1.70
Tf(K)	11.07	0.45	1.94
Vc(m3)	6.16	1.56	0.24

## 5. 감사

본 연구는 두뇌 한국 21사업의 연구 지원금에 의한 것입니다.

## 6. 참고 문헌

1. Christopher M. Bishop : "Neural Networks for Pattern Recognition", Oxford university press inc. (1995)
2. C. Bruneton, C. Hoff and P. I. Barton : *Computers Chem. Eng.* **22**(6), 735(1998)
3. Doug Wielenga, Bob Lucas and Jim Goerges : "Enterprise Miner: Applying Data Mining Techniques" ,SAS Institute Inc.(1999)
4. Gail F. Nalven : "Practical Engineering Perspectives Plant Safety", *AICHE*(1996)
5. H.J. Liaw, C.J. Chen, C.C. Yur : *J. of Loss Prevention in the Process Ind.* **14**, 371(2001)
5. Kevin G. Joback : Ph. D. Dissertation, Massachusetts Institute of Technology(1984)
6. Kevin G. Joback : *Fluid Phase Equilibria.* **185**, 45(2001)
7. Kun-Hong Lee, June Young Jung and In Beum Lee. : *Hwhak Konghak.* **31**(6), 744(1993)
8. Leonidas Constantinou and Rafiqul Gani : *AIChE Journal*, **40**, 1697(1994)
9. Reid, R. C., Prausnitz, J. M. and Sherwood, T. K.: "The Properties of Gases and Liquids", 5th ed., McGraw-Hill, New York, NY(2001).
10. Warren J. H., Leo Random, Paul v.R.S., John A.P : " AB Initio molecular orbital theory", John Wiley & Sons, Inc(1986)