



순수 성분의 물성 자료를 이용한 2성분계 혼합물의 인화점에 대한 다변량 통계 분석 및 예측

김성영 · †이범석

경희대학교 화학공학과

(2007년 5월 28일 접수, 2007년 7월 24일 채택)

Multivariate Statistical Analysis and Prediction for the Flash Points of Binary Systems Using Physical Properties of Pure Substances

Sung-Young Kim · †Bomsock Lee

Dept. of Chemical Engineering, KyungHee University

(Received 28 May 2007, Accepted 24 July 2007)

요 약

다변량 통계 분석법(Multivariate statistical analysis method)의 대표적 방법인 다중 선형 회귀법(Multiple linear regression, MLR)을 이용하여 2성분계 혼합물의 인화점을 회귀 분석하고 예측하였다. 가연성 물질의 인화점에 대한 예측은 실제 화학 공정 설계에서 화재 및 폭발 위험성을 판단하는 중요한 부분 중의 하나이다. 본 연구에서는 순수 성분의 물성 자료만을 이용하여 2성분계 혼합물의 인화점 실험 자료에 대해 다중 선형 회귀법(MLR)을 수행하였고, 이를 이용하여 새로운 혼합물에 대한 인화점을 예측하였다. 2성분계 혼합물의 인화점에 대한 MLR의 회귀 성능과 새로운 혼합물에 대한 예측 성능을 알아보기 위해, 기존의 인화점 추정 방법인 Raoult의 법칙과 Van Laar식에 의한 추정값과 비교해 보았다.

Abstract – The multivariate statistical analysis, using the multiple linear regression(MLR), have been applied to analyze and predict the flash points of binary systems. Prediction for the flash points of flammable substances is important for the examination of the fire and explosion hazards in the chemical process design. In this paper, the flash points are predicted by MLR based on the physical properties of pure substances and the experimental flash points data. The results of regression and prediction by MLR are compared with the values calculated by Raoult's law and Van Laar equation.

Key words : Flash point, Regression analysis, Multiple linear regression, MLR, Multivariate statistical analysis, Pure substances

I. 서 론

화학 반응이 복잡하고 반응 상수들에 대한 정보가 부족한 실제 화학 공정에 대해서 입출력 데이터를 기반으로 하는 실험적 모델이 사용되고 있다. 실험적 모델은 다중 선형 회귀법(multiple linear regression, MLR)과 부분 최소 사승법(partial least squares, PLS) 등의 다변량 통계 분석법에 의해 만들어지며, 복잡한 화학 공정에서 생산되는 제품의 물성 예측이나 분석에 널리 사

용되고 있다. 본 논문에서는 다변량 통계 분석법의 대표적 방법중의 하나인 다중 선형 회귀법(MLR)을 이용하여 2성분계 혼합물의 인화점을 회귀 분석하고 예측하였다.

인화점은 가연성 액체의 화재 위험성을 나타내는 지표로써, 공기 중에서 가연성 액체의 면에 증기가 발생하여 그 증기가 착화원에 접근할 경우 인화되는 액체의 최저온도로 정의된다. 이는 Fig. 1과 같이 폭발 하한(lower flammable limit, LFL)과 혼합물의 포화증기가 만나는 점에 해당하는 온도이다. 화학공정에 사용되는 가연성 물질들의 인화점 예측은 화학공정의 위험성 검

*주저자:bslee@khu.ac.kr

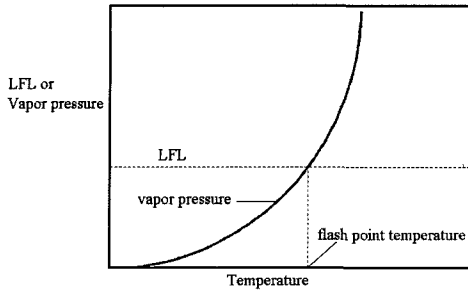


Fig. 1. Flash point by means of vapor pressure and flammable limit.

사를 위해 매우 중요하다. 이는 실제 공정에서 취급 부주의로 인한 화재 및 폭발의 위험성을 지니고 있기 때문이다. 순수한 물질의 인화점은 MSDS(Material Safety Data Sheets)나 SFPE(Society of Fire Protection Engineers) handbook[3] 등의 자료로부터 확인할 수 있지만, 2성분계 또는 3성분계이상의 다성분계 혼합물의 경우에는 조성의 변화가 다양하고 측정에 많은 시간이 요구되므로 인해 인화점에 대한 자료를 확인하기가 어렵다.

Affens[7] 등이 2성분계 탄화수소 용액의 인화점 예측에 Raoult의 법칙을 도입하였고, Throne[8]은 혼합물의 인화점을 예측하기 위해 Van Laar 식과 Clausius-Clapeyron 식을 사용하여 활동도 계수를 도입하였다. Ha[9] 등은 Raoult의 법칙과 Van Laar 식을 이용해 2성분계 인화점을 예측하고 밀폐식 장치를 이용해 측정된 실험값과 비교하였다.

Eriksson[5] 등은 PLS와 MLR 방법을 이용해 수질 오염 데이터를 분석하고 그 결과를 비교하였고, Suzuki[6] 등은 유기 화합물의 점도를 예측하기 위해 PLS와 MLR 방법을 사용하였다. 그리고 Katritzky[10] 등은 다변량 통계 분석법 중의 하나인 QSPR(quantitative structure property relationship) 분석법을 이용해 순수 물질의 인화점을 추정하는 연구를 발표하였다.

Kim[12] 등은 다변량 통계 분석법을 이용하여 특정 혼합물의 조성에 따른 인화점 실험 자료를 분석하여 모르는 조성에서의 인화점을 예측하였고, 본 논문에서는 순수 물질의 증기압, 폭발 하한계(LFL), 연소열 등의 물성 자료로 구성된 입력데이터 블록과 2성분계 인화점 실험 데이터[11]를 MLR에 적용하여 기존의 2성분계 인화점의 실험값을 회귀 분석하였고, 이를 이용하여 새로운 2성분계 혼합물에 대한 순수 성분의 물성 자료를 통해 인화점 예측을 수행하였다. 이 결과들을 기존의 Raoult의 법칙과 Van Laar 식에 의해 예측된 결과들과 비교하였다.

II. 이 론

2.1. 다변량 통계 분석법(multivariate statistical analysis)

입출력 데이터가 복잡한 화학공정의 다변량 데이터를 분석하고 물성을 예측하는데 다변량 통계 분석법은 매우 유용하다. 다변량 통계 분석법은 입출력 데이터를 동시에 분석하는 통계적 기법으로 실제 실험데이터에 의존한 실험적 모델링(empirical modeling)이다. 사용되는 화학반응식이나 반응상수들이 많아서 예측 모델의 구성이 힘든 경우, 실제 실험데이터를 기반으로 모델을 세울 수 있는 장점이 있다. 복잡한 화학공정의 경우 사용되는 화학반응식이 많고 반응상수들에 대한 정확한 정보가 부족하면 정확한 모델을 구성할 수 없는 제한이 따른다. 반면에 다변량 통계 분석법과 같은 실험적 모델의 경우는 공정에 대한 정확한 이해가 없어도 실제 공정 데이터를 기반으로 모델을 세울 수 있다. 본 논문에서는 다변량 통계분석법 중 MLR을 사용하여 2성분계 혼합물의 인화점 실험데이터[11]를 기반으로 실험적 모델을 수립하여 실험값을 회귀 분석하였고, 이를 이용해 새로운 2성분계 혼합물의 인화점을 예측하였다.

2.2. 다중 선형 회귀법(multiple linear regression, MLR)

다중 선형 회귀법(MLR)[2,13]은 입출력 변수가 각각 한개씩인 단변량 데이터에 적용하는 최소 자승법(least square method)을 입출력 변수가 각각 여러개인 다변량 데이터에 적용할 수 있도록 확장한 방법으로서, 여러개의 입력 변수와 출력 변수를 한꺼번에 분석할 수 있는 장점을 가진다.

m 개의 입력 변수 $x_i (i = 1, \dots, m)$ 와 p 개의 출력 변수 $y_j (j = 1, \dots, p)$ 와의 선형관계는 다음과 같은 선형 회귀 식으로 나타낼 수 있다.

$$y_j = \sum_{i=1}^m b_{ij}x_i + e_j \quad (j = 1, \dots, p) \quad (1)$$

여기서 b_{ij} 는 입력 변수 x_i 의 출력변수 y_j 에 대한 회귀 계수이다.

입력 변수 x_i 와 출력 변수 y_j 사이의 실험적 모델을 만들기 위해 채택한 실험 횟수를 $n(k = 1, \dots, n)$ 번이라고 한다면 k 번째 실험한 입력 변수 x_{ki} 값과 출력 변수 y_{kj} 값 사이의 선형 관계식은 다음과 같이 표시될 것이다.

$$y_{kj} = \sum_{i=1}^m b_{ij}x_{ki} + e_j \quad (k = 1, \dots, n, j = 1, \dots, p) \quad (2)$$

n 번의 실험으로 얻어진 입력력 변수들의 값을 출력 변수 데이터 블록(Y) 및 입력 변수 데이터 블록(X)으로 표시하면 (2)는 행렬 방정식으로 표시할 수 있게 된다.

$$Y = XB + e \quad (3)$$

여기서 B 는 입력 변수 데이터 블록(X) 및 출력 변수 데이터 블록(Y)을 선형 관계 짓는 계수 행렬을 나타내며, e 는 오차를 나타내는 열벡터이다. 오차 e 를 최소 ($e \approx 0$)로 만드는 계수 행렬 B 는 최소 자승법을 이용하여 다음과 같은 행렬 연산에 의하여 구해진다.

$$B = (X^T X)^{-1} X^T Y \quad (4)$$

(4)로 구해진 행렬 B 를 이용하면 MLR 모델을 만드는 데 이용되었거나 이용되지 않은 입력 변수 x_i 값에 대한 출력 변수 y_j 의 예측값을 계산할 수 있다. 이와 같이 MLR은 다른 다변량 통계 분석법과 달리 사용하기가 간편하고 변수들이 갖고 있는 의미를 이해하기 쉬운 장점을 갖고 있다.

III. 2성분계 혼합물의 인화점 예측

새로운 2성분계 혼합물의 인화점 예측을 위해 8개 (Table 1)의 서로 다른 혼합물의 조성 변화에 따른 인화점의 실제 실험 데이터[11]를 이용하였다. 이 실험 데이터와 순수 성분의 물성자료를 MLR에 적용하여 알고 있는 혼합물들의 인화점을 회귀 분석하고, 이를 이용하여 2-propanol + toluene, n-butanol + n-decane 등의 새로운 2가지 혼합물에 대한 인화점을 예측하였다.

2성분계 혼합물의 증기압이 폭발하한계(LFL)의 농도와 같은 때의 온도가 인화점이므로, 혼합물의 조성과 순수 물질의 증기압, 폭발하한계(LFL), 연소열 등이 인화점 예측에 중요한 자료들이다. 본 연구에서는 MLR

Table 1. Data for regression by MLR.

n-propanol + n-propionic acid
n-butanol + n-propionic acid
n-butanol + p-xylene
n-propanol + n-butanol
n-propanol + acetic acid
n-decane + n-octanol
n-butanol + acetic acid
M.E.K + toluene

에 적용할 입력 데이터 블록(X)을 이러한 혼합물의 조성, 순수 성분의 증기압과 관련된 Antoine constants, 폭발하한계(LFL), 연소열 등으로 구성하였고, 출력데이터(Y)에 Table 1의 실제 인화점 실험 데이터를 사용하였다. 이와 같이 구성된 입력데이터 블록(X)과 출력데이터(Y)를 MLR에 적용하여 인화점을 회귀 분석하고, 새로운 2가지 혼합물(2-propanol + toluene, n-butanol + n-decane)에 대한 새로운 입력 데이터 블록을 생성하여 인화점을 예측해 보았다.

IV. 결과 및 고찰

MLR을 이용한 회귀(regression)와 예측(prediction) 결과의 신뢰성을 확인하기 위해 Raoult의 법칙, Van Laar식에 의한 계산값과 실제 인화점의 실험값과의 평균 제곱근 오차(root square error, RSE) 값을 이용하여 비교하였다. Table 2에서는 알고 있는 8개 혼합물의 조성에 따른 인화점 실험데이터에 대한 회귀 분석 결과들을 나타내었고, Table 3에서는 이를 이용하여 새로운 2개의 혼합물에 대한 예측 결과들을 나타내었다.

MLR에 의한 회귀 결과는 Raoult의 법칙과 Van Laar 식과 같은 수식 계산에 의한 추정값보다 뛰어난 결과를 얻을 수 있다. 이는 Table 2의 평균 제곱근 오차(RSE) 값의 비교와 Fig. 2를 통해 나타내었다. 이 결과를 이용한 새로운 2가지 혼합물(2-propanol + toluene, n-butanol + n-decane)에 대한 MLR의 예측 결과는 Fig. 3과 Table 3을 통해 확인할 수 있다.

MLR에 의한 회귀 분석의 결과에서 살펴보면 Raoult의 법칙과 Van Laar식에 의한 계산값보다 개선되었으므로, 이는 회귀 분석에 사용된 동일 혼합물들의 모르는 조성에서의 인화점을 추정할 때에는 다른 수식 계산에 의한 방법들보다 뛰어난 의미를 가진다. 반면, 예측 결과는 2-propanol + toluene의 경우, Raoult의 법칙에 의한 계산값보다는 예측 성능이 떨어지지만 Van Laar 식에 의한 계산값보다는 우수한 결과를 나타내었고, n-butanol + n-decane의 경우에는 Van Laar 식에 의한 계

Table 2. Comparison of regression results by MLR (for Table 1).

For regression	Average of root square error (RSE)
Raoult's law	3.8018
Van Laar eq.	3.0991
M L R	1.4095

*RSE with the experimental flash points (°C) data

Table 3. Comparison of prediction results by MLR.

	Mole fraction		Experiment (°C)	Raoult's law	Van Laar Eq.	MLR
	X1	X2				
2-propanol + toluene	0.2	0.8	5	3.17	0.55	3.66
	0.4	0.6	7	4.43	0.44	3.91
	0.6	0.4	7	5.76	0.93	4.16
	0.7	0.3	8	6.44	1.45	4.29
	0.8	0.2	9	7.14	2.38	4.41
Average of RSE				1.8120	6.0500	3.1122
n-butanol + n-decane	0.904	0.096	29	32.73	31.17	31.282
	0.796	0.204	30	33.88	31.14	32.433
	0.505	0.495	30	37.59	31.86	35.535
	0.390	0.610	30	39.42	32.56	36.761
	0.304	0.696	32	40.96	33.48	37.678
	0.209	0.791	33	42.91	35.25	38.691
	0.095	0.905	34	45.68	39.56	39.906
Average of RSE				7.8814	2.4314	3.9653

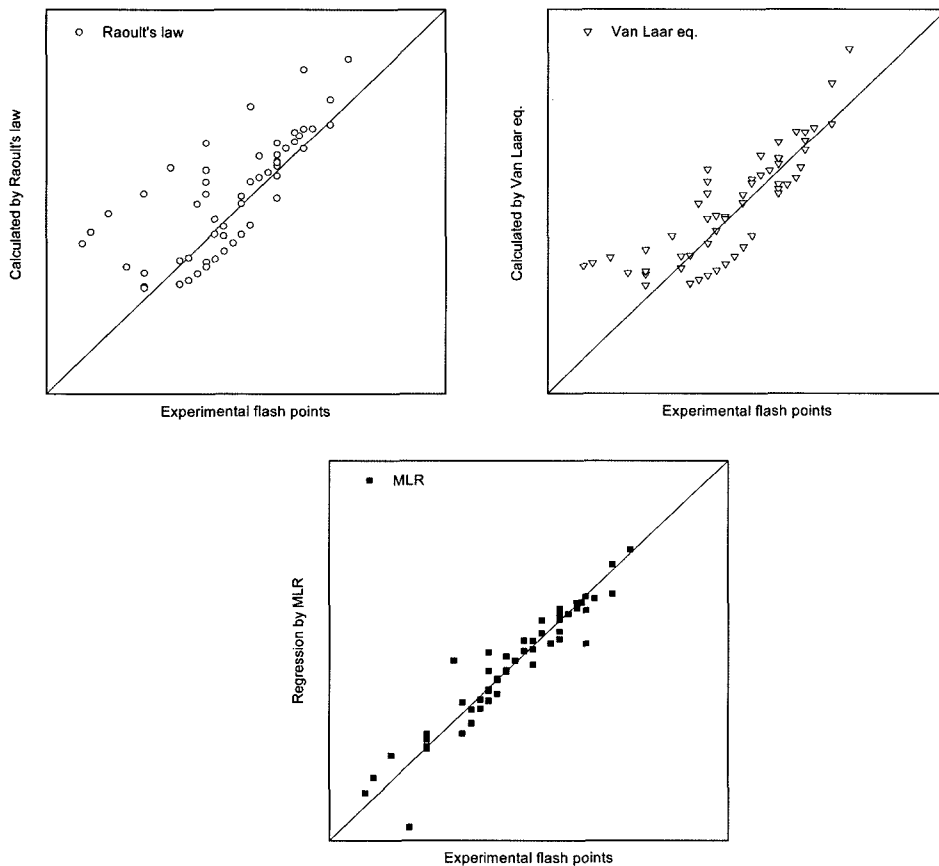


Fig. 2. Flash points (°C) calculated by Raoult's law & Van Laar eq. and regression result by MLR.

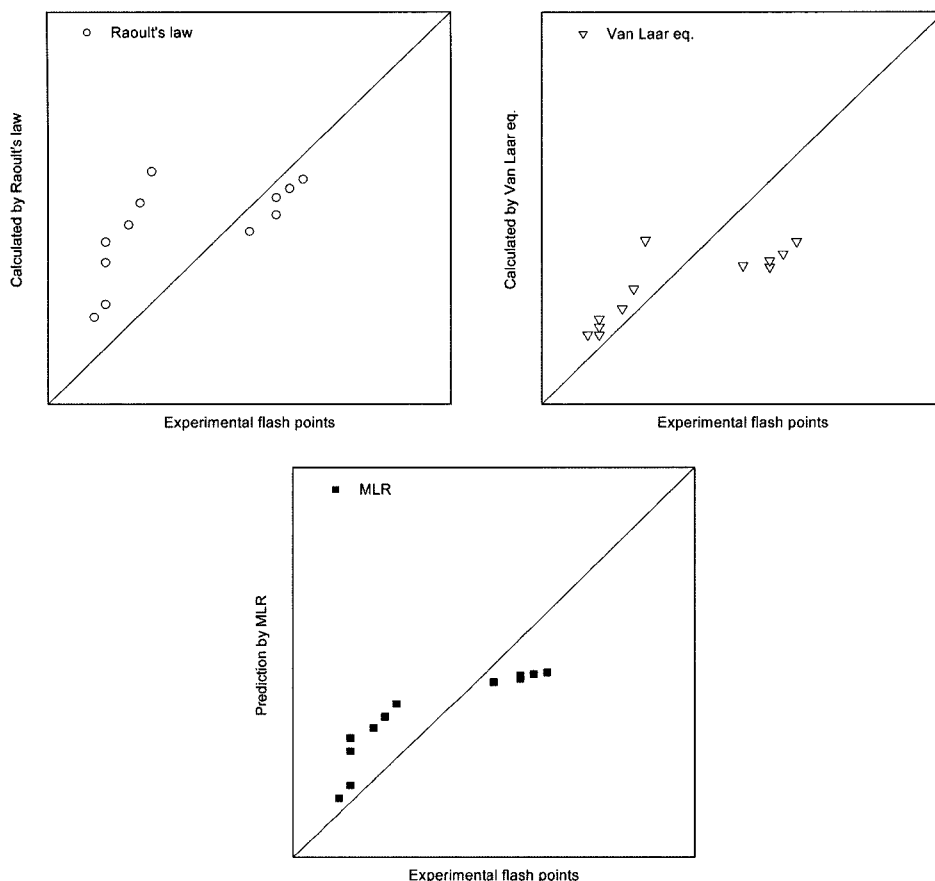


Fig. 3. Flash points (°C) calculated by Raoult's law & Van Laar Eq., and prediction results by MLR.

산값보다는 다소 떨어진 예측 성능을 나타내지만 Raoult의 법칙에 의한 계산값보다는 우수한 결과를 나타내었다. 이는 회귀 분석에 사용된 혼합물들이 아닌 새로운 혼합물의 예측에는 회귀 분석에서 만큼의 좋은 결과를 보이지 않음을 나타낸다. 그러나 MLR의 회귀 결과를 이용한 예측에서는 각각의 혼합물에 적용하는 수식 모델에 비해 간편한 수식 계산만으로 여러 혼합물들을 하나의 회귀 모델로 예측할 수 있는 장점이 있다. 새로운 2가지 혼합물 전체에 대한 인화점 예측 결과는 Fig. 3에 나타내었다.

V. 결 론

2성분계 혼합물의 인화점 실험데이터와 순수 성분의 물성 자료(Antoine constants, LFL, 연소열)를 이용해 다변량 통계 분석 방법 중 MLR을 적용하여 인화점 회귀 분석과 예측을 수행하였다. MLR을 통해 Table 1의 8가지 혼합물의 인화점 실험 데이터와 혼합물의 조성,

각 순수 물질의 Antoine constants, LFL, 연소열 등의 자료를 이용하여 회귀 분석을 실시하고 실험 데이터의 추정값을 구하였다. 이 회귀 결과를 이용해 2-propanol + toluene, n-butanol + n-decane의 새로운 혼합물들에 대한 인화점을 예측할 수 있었다. 기존의 연구에서 발표된 Raoult의 법칙과 Van Laar 식 등의 수식에 의한 인화점 계산에 비해 계산 절차가 간편하고, 물질별로 수식 모델을 적용하는 것이 아니라 대상 물질 전체를 하나의 회귀 모델로 예측할 수 있는 장점이 있다. 결과는 Fig. 2, Table 2에서와 같이 회귀에서는 기존의 Raoult의 법칙, Van Laar 식에 의한 계산값보다 뛰어나다. 예측에서는 Table 3에서와 같이 혼합물의 종류에 따라 예측 성능의 차이가 나타나고 회귀 분석에서 만큼의 좋은 결과를 보이지 않는다. 그러나 Fig. 3에서와 같이 2가지 새로운 혼합물 전체에 대한 결과를 살펴보면 기존의 Raoult의 법칙, Van Laar 식과 비슷하거나 다소 우수한 결과를 나타냄을 확인할 수 있었다. 이와 같이, MLR은 실험 데이터를 알고 있는 혼합물의 인화점에

대한 추정 능력도 다른 수식 계산 방법보다 우수하고, 실험 데이터를 알지 못하는 새로운 혼합물에 대한 인화점 또한 기존의 수식 계산에 의한 방법과 비슷한 예측 능력을 나타냄을 알 수 있다. 따라서 기존의 2성분계 혼합물의 인화점 실험데이터와 각 순수 성분의 물성 자료만을 이용한 MLR은 모르는 조성에서의 인화점 추정에서는 물론 새로운 혼합물의 인화점 예측에도 유용하다고 할 수 있다.

참고문헌

- [1] Crowl, D.A. and J.F. Louvar, *Chemical Process Safety : Fundamentals with Applications*, Prentice Hall PTR, (2002)
- [2] Geladi, P. and B.R. Kowalski, "Partial Least Squares Regression : A Tutorial", *Analytica Chimica Acta*, **185**, 1-17, (1986)
- [3] SFPE, *The SFPE Handbook of Fire Protection Engineering*, 2nd ed., Society of Fire Protection Engineers, Boston, (1995)
- [4] Gmehling, J., U. Onken and W. Arlt, *Vapor-Liquid Equilibrium Data Collection*, Vol. 1, Part 1~Part 7, Deutsche Gesellschaft für Chemisches Apparatewesen (DECHEMA), (1980)
- [5] Eriksson, J., J.L.M. Hermens, E. Johansson, H.J.M. Verhaar and S. Wold, "Multivariate Analysis Aquatic Toxicity Data with PLS", *Aquatic Science*, **57**, 217, (1995)
- [6] Suzuki, T., K. Ohtaguchi and K. Koide, "Computer-Assisted Approach to Develop a New Prediction Method of Liquid Viscosity of Organic Compounds", *Computers and Chemical Engineerings*, **20**(2), 161-173, (1996)
- [7] Affens, W.A. and G.W. McLaren, "Flammability Properties of Hydrocarbon Solutions in Air", *J. of Chem. Eng. Data*, **17**(4), 482-488, (1972)
- [8] Thome, P.F., "Flash Point of Mixtures of Flammable and Non-flammable Liquids", *Fire and Materials*, **1**, 134-140, (1977)
- [9] Ha, D.M., S.J. Lee, Y.C. Choi and H.J. Oh, "Measurement of Flash Points of Binary Systems by Using Closed Cup Tester", *HWAHAK KONGHAK*, **41**(2), 186-191, (2003)
- [10] Katritzky, A.R., R. Petrukhin, R. Jain and M. Karelson, "QSPR Analysis of Flash points", *J. Chem. Inf. Comput. Sci.*, **41**, 1521-1530, (2001)
- [11] 최용찬, "밀폐식 장치를 이용한 혼합물의 인화점 측정 및 예측", 공학석사 학위논문, 세명대학교, (2003).
- [12] Kim, S.Y., B.S. Lee, C.B. Chung and S.H. Choi, "Prediction of Flash Point of Binary Systems by Using Multivariate Statistical Analysis", *Journal of the Korean Institute of Gas*, **10**(4), 29-33, (2006)
- [13] Kim, S.Y., B.S. Lee, C.B. Chung and S.H. Choi, "Multivariate Statistical Analysis Approach to Predict the Reactor Properties and the Product Quality of a Direct Esterification Reactor for PET Synthesis", *J. of Control, Automation and Systems Engineering*, **11**(6), 550-557, (2005)