Research Reports of the Research Institute
of Industrial Safety, RIIS-SRR-90, 1990
UDC 519.654:541.126:541.65

化学構造からの熱危険性データの予測*
ーベンゼン単環誘導体ー

藤本康弘**, 安藤隆之**,
板垣晴彦**, 森崎　繁**

Prediction of Thermal Hazard Data of Benzene Derivatives by Multivariate Statistical Analysis on Their Chemical Structure*

by Yasuhiro FUJIMOTO**, Takayuki ANDO**,
Haruhiko ITAGAKI** and Shigeru MORISAKI**

Abstract: Evaluation of chemical hazards is important in the chemical industries because explosions or fires occasionally occur in the batch processes where fine-chemicals including pharmaceuticals or functional resins are produced. If the evaluation of these chemical hazards is possible by easy calculation, this may save chemical industries' money and time.

Thermal hazards of chemical substances must be evaluated from the two aspect; sensitivity, which shows how easily chemicals may decompose or ignite, and severity, which shows how much or how fast energy is liberated when chemicals decompose. One of typical methods for evaluation of thermal hazards is DSC, which gives information on sensitivity (exo-thermic onset-temperature) and severity (decomposition heat) at the same time with easy operation. This report consists of results of multivariate regression analysis and discrimination analysis between DSC-data and chemical structures of benzene derivatives (307 samples). If these results could predict chemical hazards accurately enough, they would be used for direct evaluation of thermal hazards from chemical structures which are easily known without any experiments.

(1) Regression Analysis

The following three sets were selected as suitable variables for pre-regression analysis in mono-substituted nitrobenzenes (51 samples) for the prediction of decomposition heat (kcal/mol).

SET 1; Types of functional groups

SET 2; Atoms in functional groups

SET 3; Chemical bonds in functional groups

From the results of calculating of the coefficient of determination (R^2) and the correlation coefficients, SET 3 was found out to be the most suitable model. The regression analysis using the variables in SET 3 in benzene derivatives (307 samples) gave relatively good result, showing R^2 of 73.5%. However, another set of variables including influence of interaction between functional groups may be needed to improve the accuracy of this prediction. For the prediction of exothermic onset-temperature, SET 1 and SET 3 were also selected as suitable sets of variables in the same

*第 22 回（1989 年 12 月 8 日）及び第 23 回（1990 年 12 月 7 日）安全工学研究発表会において一部発表[1,2)]
**化学研究部, Chemical Safety Research Division

way. But the both models didn't show good prediction. This reason may due to the fact that exothermic onset-temperature is primarily influenced by minimum bonding energy in chemical substances.

(2) Discrimination Analysis

Identification of thermal hazard rank of each di-substituted benzenes (128 samples) was carried out using discrimination technique on decomposition heat (cal/g). The kind of chemical bonds in functional groups was adopted as variables and the correct discriminate ratio was over 90%.

However, our chemical knowledge is not enough to explain the meaning of the coefficients in the discrimination analysis. And more, it was difficult to decided the rank of chemicals which had the values of decomposition heat near the border (ex. 200 cal/g). It may be effective to apply fuzzy theory on this discrimination analysis to solve the above problem.

Key Words: Thermal Hazard, Multivariate Analysis, (Regression, Discrimination), DSC, Exothermic Onset-temperature, Decomposition Heat

1. 緒　言

化学工場等で新規の化学物質を扱う場合などに，取り扱いの不備により爆発・火災事故が発生することがある。これらの事故を防止する為には，その化学物質の危険性（感度，威力）がどの程度であるかを取り扱う前に充分に把握しておく事が重要である。

そこで当該化学物質を使用する前にその化学物質の危険性を評価しておく事が望ましいわけだが，その危険性判定の方法としては，試験せずに判定する事前評価と実際に試験を行う方法の二つがある。(Fig. 1)。

この両者の方法を比較した場合は，勿論実際に試験を行って危険性を判定する方がより正確な値を得られるのは明らかである。その試験方法にはスクリーニング試験，標準試験，大規模／実規模試験といったものがあり，この中では，大規模／実規模試験が実際の操作中の危険性を最も的確に判定できる。

しかし，実際に大規模／実規模試験を行うには危険が伴うので，測定装置，実験場所の確保等の多大な設備投資が必要となる。したがって中小企業等では自らそのような設備を準備するのは困難であり，公的な試験所等に依頼するしかない。そして，試験結果を得るまでにも長い時間を要する。この結果，取り扱う化学物質全てについて実際に試験を行って危険性を評価するのは非常に難い。これらの点は他の試験（標準試験，スクリーニング試験）でも程度の差こそあれ同じ様に問題となる。

そこで何らかの事前評価によって化学物質のスクリーニングを行い危険性の等級分けをして，危険性が高いと推定される化学物質を，優先的に実際に試験を行うことが有効な手段となる。

このような事前評価の手法は既存化学物質の危険性のチェックのみならず，新規の化学物質の開発時の危険性の事前評価にも有効な手法となり，研究開発の効率化に貢献するものと期待される。

この様な事前評価の方法としては，Fig. 2 にあげた様に二つのアプローチが考えられる。まず理論的計算（熱力学的計算）によるものでは，例えば反応熱を求めようとすれば原系と生成系の生成熱を知らねばならない。その場合，まず分解生成物（生成系）の構造を知る必要がある。もし，実測できなければ推定しなければならない。通常は空気中での熱分解の場合は最終酸化物，例えば炭素原子は二酸化炭素に，水素原子は水に完全に酸化されると考える。そして，原系，生成系それぞれの生成熱が実測されていなければそれぞれの生成熱も計算で求めなければならないが，その計算には分子軌道法による計算や Benson の二次加成則を利用したものが知られている。

この方法は単純な熱分解でその分解の挙動がよく知られていれば正確な値が得られる事が期待されるが，今日我々が扱っている，あるいは扱おうとする化学物質の多くはその構造が複雑であり，実際の作業中に生じうる熱分解生成物は理論で推定できる生成物とは全く異なる事も少なくない。

そのような場合には，得られてくる計算結果が実

1.事前評価
- 理論的計算
- 統計的分析

2.試験
- スクリーニング試験
- 標準試験
- 大規模試験

Fig. 1　Tests for hazard evaluation of reactive chemical substances.
化学物質の危険性評価

1.理論的計算　→　熱力学的計算
原系、生成系の生成熱
Ｂｅｎｓｏｎ二次加成性則
分子軌道法

2.統計的分析　→　多変量解析
同一条件下でのデータの蓄積

Fig. 2　Pre-tests for hazard evaluation of reactive chemical substances.
化学物質の危険性の事前評価

際の発熱量とはまったく異なる値になってしまう（例えば最終酸化物を仮定して計算しても，現実の条件下で実際には途中で分解が終了すれば，発熱量は計算値よりかなり小さくなる）。勿論そういった場合でもこの計算値はその化学物質の持つ危険性の最大のポテンシャルを教えてくれるので，そのデータが重要であることは論をまたない。

もう一つの事前評価の方法としては，実測値を数多くの化学物質について測定しておいて，化学構造等との相関を統計的に解析する方法が考えられる[3,4]。

この方法では，あらかじめ色々な官能基をもつ数多くの化学物質についての同一条件での危険性のデータ（例えば分解熱，発熱開始温度，圧力発生速度等）が蓄積されていることが必要となる。そしてそれらのデータを基に，構造を特徴づける因子に対しての予測式を求める事となる。

これらの実測値はすでにかなりの測定誤差が含まれており，また因子の選び方でも予測式自体に統計誤差が含まれる。したがって，ここで得られる予測値の絶対値にはかなりの誤差が含まれる事は避けられない。

それでも理論的な計算と違い，実際の分解による危険性のデータを基にしているので，異なる官能基間の相対的な危険性の大小の判定等は，理論的計算よりも実際の状況をより的確に評価することができると考えられる。

幸い，当化学研究部では数年前よりDSCを用いて同一条件での多数の有機化合物の熱分解データを継続して測定しており，これまでに800種類以上の化合物についてのデータが得られている。これらのデータは，産業安全研究所安全資料として2冊にわけて発表されている[5-7]。

そこで，この報告書では，化学構造からそのエネルギー的な危険性を予測する手法の開発を目的として，上記安全資料の中からベンゼン単環誘導体についてのDSCの測定結果（発熱量，発熱開始温度）をピックアップして，化学構造との関係を多変量解析により検討を行っている。

2. 解　析

2.1　対象とする化合物

化学物質の危険性を予測しようとする場合は，予測式の精度を考えて化学構造の類似したグループごとに検討を行うのが望ましいと考えられるが，実際には試験データの不足により，どうしても充分な検討ができないグループがでてくる。

したがって充分なデータの蓄積が行われるまでは予測式の精度は落ちるが，化学物質の構造の類似性を余り厳しくせず，統計的な分析を行うのに充分なデータが蓄積されているグループを対象とせざるを得ない。ここでは比較的データが多く推定されている化学物質として，基本骨格にベンゼン環を一つだけ持つベンゼン単環誘導体のグループ（側鎖の官能基の数，種類には特に限定を設けない）を対象として解析を行った。

2.2　因子（説明変数）

化学物質の構造と危険性の関係を分析する場合，予測式に含める説明因子としては二つの大きな因子が考えられる。

Table 1　Classification of chemical hazards and their evaluation indications.
化学物質の危険性分類とその指標

危険性＼指標	感度	威力
熱発火・分解危険性	発熱開始温度 分解開始温度 発火温度	分解熱 燃焼熱 断熱温度上昇 最大発生圧力 圧力発生速度 分解・燃焼速度
着火・燃焼危険性	引火点 燃焼点 最小発火エネルギー	燃焼速度 燃焼熱 圧力発生速度
衝撃危険性	打撃感度 摩擦感度	分解熱 爆発熱 分解速度 圧力発生速度
反応危険性	発熱開始温度 発火温度他	反応熱 燃焼熱 温度上昇速度 圧力発生速度

その一つは原子の種類，官能基の種類や数など構造式から直接知る事ができるもので，もう一つは官能基の持つ電子吸引性，電子供与性，例えばオルト効果として知られる空間を通しての水素結合の影響等，官能基の電子的な性質に由来するものである。

この両者は相補うもので，どちらか一方だけでその化学物質の構造の特徴を完全に記述することは当然不可能である。しかし，ここで後者の電子的な性質の効果を，対象とする化学物質のグループ内の化合物すべてについて知る事は，現実問題としては非常に難しく，容易に入手できるデータはごく限られたものとなる。

統計的な解析においては，データが不充分では意味のある解析を行うのは難しくここでは後者のデータは省き，構造式から容易に判断できる前者の因子のみを用いて予測式を求める事にした。将来的には電子的な因子を含めた解析に発展させる予定である。

2.3　予測指標

化学物質の危険性には大きく分けて４つに分類できると考えられる。それは熱発火・分解危険性，着火・燃焼危険性，衝撃危険性，反応危険性であり，それぞれの危険性について感度（種々のエネルギーが与えられた時の分解，爆発，反応の起こしやすさ）と威力（分解，爆発，反応に伴うエネルギーの発生量及び発生速度）の二つの視点から見ていく必要がある。これら危険性と視点の面からみた具体的な（測定可能な）指標が Table 1 にまとめてある。

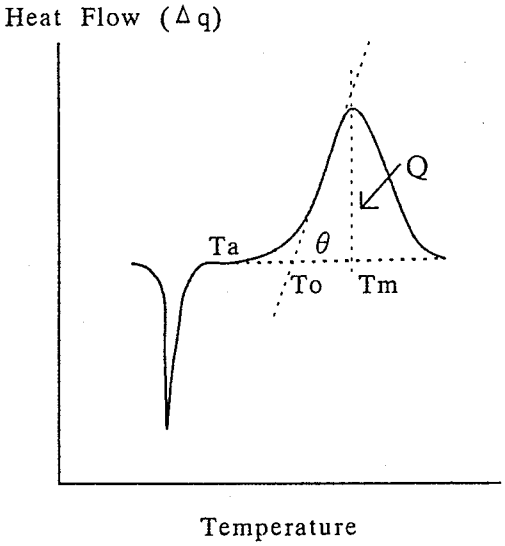

Fig. 3　Thermal hazard informations from DSC curve.
DSC から得られる熱危険性の情報

これらの危険性の中で，ここでは当化学研究部において DSC によるデータの蓄積がかなり行われている熱分解危険性に注目した。

DSC は，感度を示す指標（発熱開始温度等）と威力を示す指標（発熱量，最大発熱加速度等）を同時に測定でき，また熱危険性の標準スクリーニング試験として広く用いられており，消防法の第五類の試験法にも採用されているなど熱分解危険性を測定する代表的な試験法である。

DSC の測定データから読み取り可能な情報としては，発熱開始温度Ta, To，発熱ピーク温度Tm，発熱量Q，最大発熱加速度 $\tan\theta$等があげられる (Fig. 3)。ここでは感度指標として発熱開始温度To，威力指標として発熱量Q を選択した。

2.4　危険性評価

危険性評価方法の基本的な考え方については，「化学物質の危険性予測システムの開発研究」[3]に詳しく述べられているのでここでは概略を述べる。

化学物質の危険性の予測・評価する為には設定された各指標を総合的に判断する事が望ましい。まず今回取り上げた二つの予測指標（発熱開始温度，発熱量）と危険性との関係は定性的には次のようになると考えられる。

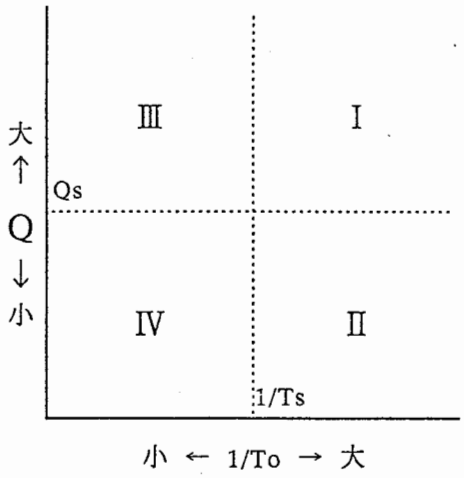

Fig. 4　Thermal hazard classification by To and Q.
発熱開始温度Toと発熱量Qによる危険性の分類

(1) 危険性の高い物質
発熱開始温度が低く，かつ，発熱量が大きい物質
(2) 危険性の低い物質
発熱開始温度が高く，かつ，発熱量が小さい物質

そこで，次にこれらの予測指標を総合的に評価する方法が必要となるが，その方法として，以下の二通りの方法を述べる。

まず，第一の方法は，各予測指標毎に危険性のランクづけをおこなって各指標を総合したランクに分ける方法である。例えば発熱開始温度と発熱量とがそれぞれ2つにランク分けすることができれば総合的には4つのランクに分類することができる。ランクの予測は，判別分析の手法を用いることで判定できる。危険性はランクIが最も大きく，ランクIVが最も小さい。ランクII, IIIはその中間である。

この結果，直接，危険，非危険の判定が精度が良くできれば化学物質の法規制等の基準にも使うことができる (Fig. 4)。

この方法では，各指標のランク分けの基準の値 ($1/Ts, Qs$) をいくらかにするかということが大切になる。また，それ以上に問題なのは判別式を求める為の測定値にはかなりの誤差が含まれているので，基準値付近の化合物はごく僅かの変動で危険物になったり，非危険物になったりするということである。この点についてはその改良法として最近ではファジー理論を応用した解析方法が報告されている。

次の方法は，各予測指標を総合化した新たな総合指標を設定し，その指標を用いて危険性を予測，評価する方法である。

Table 2　Conditions for DSC experiment.
DSC 測定条件

測定装置	：	熱流束方式加圧型 DSC
昇温速度	：	10°C/min
試料量	：	1 ～ 2mg
試料容器	：	ピンホール容器（アルミ製）
雰囲気	：	アルゴン加圧（34kg/cm^{2}G）
測定化合物	：	ベンゼン単環誘導体

総合指標の設定方法が問題であるが例えば一つの例として次のような式が挙げられる。

$$\text{総合指標} = W_{To}(\frac{1}{To}) + W_Q Q \qquad (1)$$

W: 各指標の重み，To: 発熱開始温度，Q: 発熱量

この方法では各指標の値を重回帰分析であらかじめその化学構造から予測する必要がある。また，各指標の重みをどのように決定するかも問題である。しかし，数値で危険性が評価されるので，ランクづけの方法と比べて，その化合物がどの程度危険なのかがある程度定量的に評価できる。実際には工場等で各種の化合物を扱っている場合，この化合物は以前の化合物と比較してどの程度危険になっているかという目安として利用できると思われる。ただし，危険物かどうかの判定に利用する時は，先のランクづけの方法と同様に基準値付近の化合物の判定に誤差が生じる。ただ，境界付近かどうかは総合指標で判定できるので不確かなものは例えば準危険物として取り扱うなど柔軟な対応が可能となる。

2.5　解析データ

解析の為のデータとしては，当研究所で測定したDSC 測定結果をまとめた安全資料（RIIS-SD-87, No.1 および RIIS-SD-89, 試験総数 843 件）の中からベンゼン単環誘導体についての発熱開始温度Toと発熱量Qのデータを用いた。試験総数は 307 件であり，その測定条件を Table 2 に記す。

この解析に用いたベンゼン単環誘導体の集合について，その特徴を明らかにするために因子としての化合物官能基の数，種類の分布 (Fig. 5, Fig. 6)，予測指標としての発熱量Q，発熱開始温度Toの分布とその相関図 (Fig. 7, Fig. 8, Fig.9) を示す。

それぞれの図から特徴をみると，Fig. 5 より官能

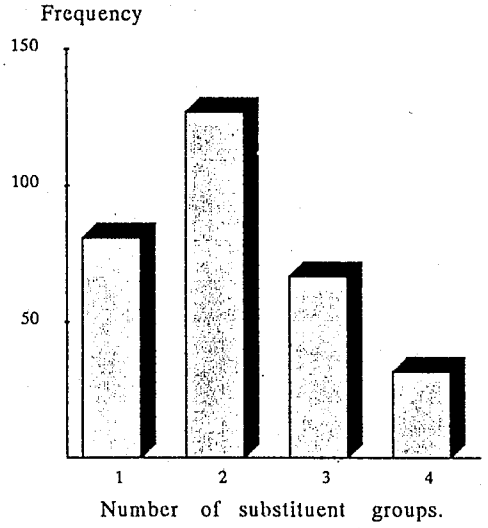

Fig. 5　Frequency distribution of number of substituent groups.
ベンゼン単環化合物中の置換基の数の分布

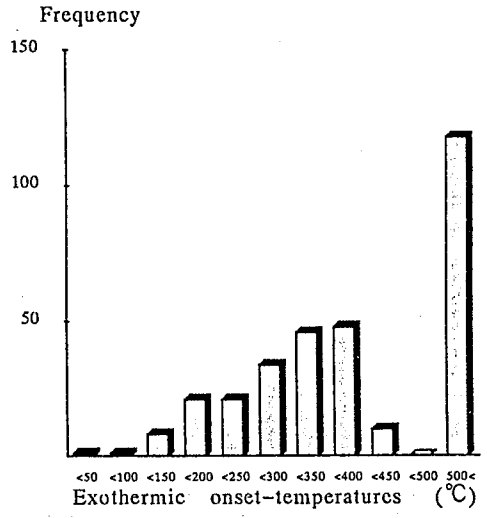

Fig. 7　Frequency distribution of exothermic onset-temperatures.
発熱開始温度の分布

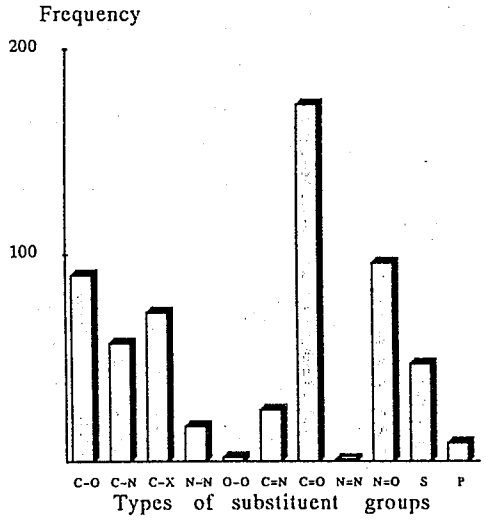

Fig. 6　Frequency distribution of types of substituent groups.
ベンゼン単環化合物中の置換基の種類別の分布

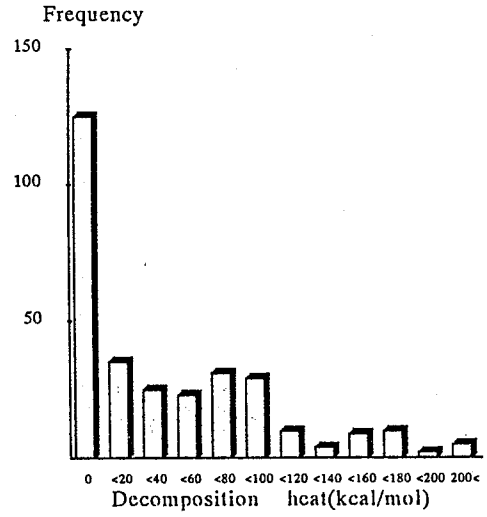

Fig. 8　Frequency distribution of decomposition heat.
発熱量の分布

基の数は 4 個以下であり，また Fig. 6 から，官能基の中では酸素原子を含むアルコール，カルボニル，ニトロ，ニトロソ基の数が多いことがわかる。また，Fig. 9 をみれば，予測指標の発熱量と発熱開始温度との間の相関係数は小さく，各々が独立の因子であることが確認できる。

2.6　分析手法

解析のソフトは，芳賀（東京理科大学・工学部）によるパーソナルコンピューター用の対話型データ解析プログラム micro-CDA を用い，化学構造を特徴づける説明変数から発熱量等を予測する式を求めた[8]。解析は，最初に適当な説明変数群のモデルを用いて重回帰分析および判別分析を行いその残差を調べ，そしてより最適なモデルを探索して分析を繰り返すという手順を踏んだ。

それぞれの解析は，変数を増減しながら最適な式を求める逐次選択法を用いたが，その内容については付録にまとめてある[8-10]。

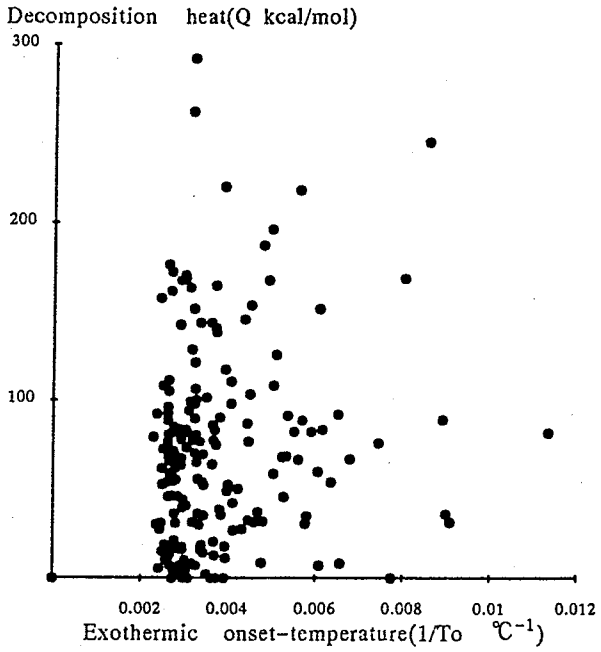

Fig. 9 Relationship between To and Q.
発熱開始温度と発熱量との相関図

3. 重回帰分析の結果

3.1 重回帰分析に用いる予測指標と説明変数

予測指標としての発熱開始温度，発熱量の単位について，特に発熱量は，現実の取り扱い時の危険性からみれば単位重量当たりで考察するのが実際的であるが，分子構造の特徴を説明変数とするからには1モル当たりの発熱量の方がより良い予測式が得られると考えられる。

また，発熱開始温度については基準点をどうするか（摂氏温度でよいのか，絶対温度が良いのか，あるいは消防法の爆発性評価で用いられている様に室温25℃を基準にするのか），あるいは，変数の変換が必要ではないか（例えば逆数変換あるいは対数変換）といった問題が在る。ここでは取り敢えず摂氏温度を基準として，また特に変数変換は行わないで解析を行うこととする。

次に，説明変数のモデルについて考える。説明変数としては，化合物の構造式から容易にその値がわかるものでないと，対象とする化学物質のグループ内の化合物全部についての値を得る事が困難になる。ここでは以下の (a) ～ (d) の説明変数について検討を行った。

Table 3 Classification of functional groups.
官能基の分類

官能基の種類 (a.0)	原子 (a.1)	結合 (a.2)
アルキル	炭素原子	炭素-炭素
カルボニル	酸素原子	炭素-酸素
(エステル，カルボン酸)	窒素原子	炭素-窒素
(アミド)	硫黄原子	窒素-窒素
アミン	ハロゲン原子	窒素-酸素
ヒドラジド	リン原子	酸素-リン
等	等	等

(a) 官能基の構造

化合物の構造の一次元的な表記法には，例えばIUPACの命名法やCASの命名法等が有名である。しかし，それらの表記は化合物の検索には便利でも，今回の目的である熱分解データを予測するための指標としての構造の表記方法としては必ずしも適切とは思えない。

不安定な化合物は分子内に固有の結合グループを有しており，それらのグループがあるエネルギーレベルに達した時にそこから反応が開始，順次連鎖的に分解が進行すると考えられる。その結合グループとしては，Brethrickの危険な反応性グループ，あるいはBenson二次加成性則を計算する際のグループ分けが参考になる。また，前記した当研究所のデータ集は一般的な官能基でグループ分けしている。

ここでは前段階としてデータ集の分類方法を参考に，ベンゼン環に結合している結合グループを官能基の種類別に分類した (Table 3)。

(a.0) 結合グループ中の官能基の種類

ただし，この分類だけでは，一つの結合グループ中に複数の官能基を持つ場合の処理が厄介になる。そこで他の情報として原子の種類，結合の種類という因子を追加した。これは熱分解においては，各原子間の結合の解裂を伴うことから，官能基の種類名と併用すればより的確に各結合グループ毎の熱分解データの説明が可能と考えた為である。

(a.1) 結合グループ中の原子の種類

(a.2) 結合グループ中の結合の種類

構造以外に容易に得られる情報として次の変数も同時に検討した。

(b) ベンゼン環との結合原子，結合位置

(c) 分子量

(d) 酸素バランス

Table 4　A result of multivariate regression analysis for decomposition heat by kinds of functional groups. (mono-substituted nitrobenzene)
発熱量の予測 1 ー官能基の種類（一置換ニトロベンゼン）

Variables	Coefficients	Std.Dev.
Const.	+62.8	6.3
−CH=CH−(Alkene)	+58.1	22.2
−(C=O)−(Carbonyl)	+12.8	11.9
−(C=O)$NHNH_2$	+48.0	28.9
−N−(Amine)	+17.5	10.8
−O−(Ether)	−7.2	12.4
O−P(Phosphate)	+133.5	61.3

今回の解析ではパソコンを用いている為，メモリーの制約で手持ちの解析データすべてについて，上記のすべての変数を候補とした解析は不可能であった。そこで，解析データの部分集合である一置換ニトロベンゼン（試験数 51 件）等のデータを利用して解析を進めて，最適と思われる説明変数群を選び，次にその説明変数群を用いて，最終的にベンゼン単環化合物（ 307 件）の解析を行った。

3.2　発熱量の予測

3.2.1　モデルの検討

まず，発熱量を予測するのに最適なモデル（説明変数）の検討を行った結果を示す。

最初に，一置換ニトロベンゼンを対象にして，下記の各モデルでモル単位の発熱量の予測式を求めた。それぞれのモデルは前節 (b) ～ (d) の説明変数をすべて含み，(a) の説明変数の取り方を変えたものである。結果については，回帰係数が零でないものを Table 4 ～ 6 にのせてある。またR^2は以下の通りである。

1. 官能基の種類を中心としたモデル
   (a.0)+(b)+(c)+(d)
   (Table 4, R^2=76.6%)
2. モデル 1 に官能基中の原子の種類別の数を追加したモデル
   (a.0)+**(a.1)**+(b)+(c)+(d)
   (Table 5, R^2=87.1%)
3. モデル 1 に官能基中の結合別の数を追加したモデル
   (a.0)+**(a.2)**+(b)+(c)+(d)
   (Table 6, R^2=82.4%)

なお，この中で，(a.0) の説明変数だけは該当する官能基の有無を質的変数として，数量化理論第一を利用して回帰分析を行っている。これは，この説明変数（官能基の種類）には，一置換ニトロベンゼンを対象とした場合，その数には何の意味もなく，存在の有無だけが重要であると考えたからである。

Table 5　A result of multivariate regression analysis for decomposition heat by kinds of functional groups and their atoms. (mono-substituted nitrobenzene)
発熱量の予測 2 ー官能基および原子の種類（一置換ニトロベンゼン）

Variables	Coefficient	Std.Dev.
Functional Group		
−(C=O)− (Carbonyl)	−13.3	6.1
O−P(Phosphate)	+133.9	16.5
Atoms in Functional Group		
Unsaturated Carbon	+12.4	3.4
Saturated Carbon	−12.8	4.3
Total Oxygene	−12.5	4.8
Unsaturated Sulfur	−83.5	26.7
Saturated Sulfur	−41.8	22.8
Molecular Weight	+0.70	0.09

さて，この中で最も良いR^2の値を示したのは，モデル 2. (Table 5) であった。この結果を用いて発熱量の予測をするためには，まず，予測しようとする一置換ニトロベンゼンのニトロ基以外の官能基中に Table 5 に載っている原子等がどれだけあるかを数え，回帰係数をその数に乗じて和を取れば良い。

例えば，モデル 2. の結果 (Table 5) を用いた予測例として，官能基として-(C=O)-O-CH_2-SH を持つ，ニトロベンゼン（分子量 213 ）を考える。するとこの分子の発熱量 (kcal/mol) の予測値は以下の様に計算される。

$$
\begin{aligned}
Q = {} & 213 \times 0.7 \text{（分子量 213）} \\
& + 1 \times -13.3 \text{（カルボニル有り）} \\
& + 1 \times 12.4 \text{（不飽和炭素 1 個）} \\
& + 1 \times -12.8 \text{（飽和炭素 1 個）} \\
& + 2 \times -12.5 \text{（酸素 2 個）} \\
& + 1 \times -41.8 \text{（飽和イオウ 1 個）} \\
& = 68.6 (\mathrm{kcal/mol})
\end{aligned}
$$

しかし，この回帰分析の結果 (Table 5)，すなわち

Table 6 A result of multivariate regression analysis for decompositon heat by functional groups and their bonding. (mono-substituted nitrobenzene)
発熱量の予測 3 －結合の種類
(一置換ニトロベンゼン)

variables	Coefficient	Std.Dev.
Const.	+65.7	3.0
C=C−(C=O)OH	+55.3	10.3
C−N	+15.2	6.9
C−X	+23.2	12.5
N−N	+23.2	8.7
O−P	+32.6	4.2
P−Cl	+81.9	8.9

Table 7 Estimate decomposition heats(kcal/mol) from a result fo multivariate regression analysis on kinds of bonding.
(mono-substituted nitrobenzene)
発熱量の予測値 (kcal/mol) －結合の種類
(一置換ニトロベンゼン)

Chemicals	Functional groups	Obs.	Pred.
p-nitrotoluene	$-CH_3$	54.1	65.7
p-nitrobenzoic acid	−(C=O)OH	67.8	65.7
p-nitrobenzamide	$-(C{=}O)NH_2$	76.7	80.9
p-nitrophenylhydrazide	$-NHNH_2$	91.1	88.9
parathion	$-OP(OC_2H_5)_2S$	140.0	163.5

回帰係数の大きさや符号の化学的な意味が余りはっきりせず，また予測式も多少複雑である。

それに対して，事前の考察では最も発熱量に相関がありそうな官能基中の結合を変数としたモデル (Table 6) による予測式は，上記の予測式と比べR^2の値が若干劣るものの，その予測式は比較的単純で分かり易い。

この結果を用いて任意のニトロベンゼン誘導体の発熱量の予測方法は上と同様に行う。例えば，その予測例として，先程と同じ化合物を考える。この化合物の場合，官能基中には Table 6 にある結合を含まないので，その発熱量の予測値は定数項のみとなる。即ちQ=65.7(kcal/mol) である。他の例として，ニトロ基以外の官能基として，$-CCl_2-NH-NH_2$を持つ化合物を考えると，官能基中に$C-X$ (Xはハロゲン) 結合が 2 つあり，$C-N$結合が 1 つ，$N-N$結合が 1 つ有るので，次の式で予測値を求める事ができる。

Table 8 A result of multivariate regression analysis on kinds of bonding.
(di-substituted benzene)
発熱量の予測 4 －結合の種類
(二置換ベンゼン)

Variables	Coefficient	Std. Dev.
C=C−(C=O)OH	+49.0	13.4
C−O	+4.6	2.9
C−Cl	+47.3	6.4
C−Br	+40.6	6.4
N−N	+33.4	7.6
N−O	+70.2	18.5
O=N=O	+56.5	3.8
N=O	+27.6	18.9
O−O	+76.6	26.3
O−P	+32.0	6.4
O=P	+53.0	22.7
S−Cl	+23.7	13.1
P−Cl	+54.8	8.3

Table 9 A result of multivariate regression analysis on kinds of bonding.
(benzene derivatives)
発熱量の予測 5 －結合の種類
(ベンゼン単環誘導体)

Variables	Coefficient	Std. Dev.
C=C−(C=O)OH	+37.3	14.8
C−O	+5.4	1.9
C−Cl	+59.1	6.8
C−Br	+40.6	7.3
N−N	+34.8	7.8
N=N=N	+87.9	18.8
N−O	+57.5	13.5
O=N=O	+67.4	2.8
N=O	+28.2	13.9
O−O	+69.2	23.1
O−P	+23.7	5.2
O=P	+61.9	26.6
S−N	+25.5	12.0
(S=O)C=C	+48.5	32.6
P−Cl	+52.8	10.3

$$Q(\mathrm{kcal/mol}) = 65.7+2\times 23.2+15.2+23.2 = 104.1$$

Table 7 にこの予測式による予測値と実測値の例をごく一部示す。このような簡単な予測式にしては比較的実測値と予測値とは良く一致している。

ここで，上記の結合の種類を説明変数としたモデルを用いて予測対象を二置換ベンゼンにして，解析を行った結果を示す (Table 8, R^2=75.4%) 。R^2の

Table 10　Functional groups which have large errors in composition heats between observed and predicted values.
予測値の残差が大きな化学物質の官能基

Functional Groups				Errors
Cl	Cl	Cl	OH	138.0
Cl	Cl	Cl	OCH_2COOH	84.9
Cl	Cl	Cl	$O(CO)OC(CH_3)_3$	104.6
Cl	Cl	Cl	NH_2	78.8
Cl	Cl	COCl	H	85.9
Cl	Cl	OCH_2COOH	H	111.0
Cl	Cl	NHC_2H_5	NH_2	82.0
Cl	Cl	$NHCOOCH_3$	H	102.6
Cl	Cl	NO_2	NO_2	104.8
Cl	COOH	SO_2NHNH_2	H	107.7
Cl	COCl	H	H	91.9
Br	Br	Br	NH_2	92.1
I	I	I	OH	80.3
CF_3	$N(C_3H_7)_3$	NO_2	NO_2	85.2
CH_3	NO_2	NO_2	NO_2	89.8
H	H	NO_2	NO_2	–82.3

Table 11　A result of multivariate regression analysis fro decomposition heat by kinds of bonding and halogen bonded to benzene ring. (benzene derivatives)
発熱量の予測 6
結合の種類＋環結合のハロゲン数
（ベンゼン単環誘導体）

Variables	Coefficient	Std. Dev.
Bonding		
C=C–(C=O)OH	+39.6	13.3
C–O	+4.3	1.7
C–Cl	+51.2	6.2
C–Br	+39.7	6.5
N–N	+32.8	7.0
N=N=N	+74.5	17.0
N–O	+58.8	12.1
O=N=O	+66.1	2.5
N=O	+30.5	12.5
O–O	+59.9	20.8
O–P	+20.0	4.7
O=P	+56.0	23.9
S–N	+18.0	10.8
(S=O)C=C	+48.5	29.2
P–Cl	+54.7	9.3
Halogen bonded to benzene ring		
ϕ–X	+20.8	2.3

値は低くなり，一置換ニトロベンゼンの場合と比べ変数の数が多くなるが，それぞれの係数の大きさはある程度納得できるものであり，一置換ニトロベンゼンのデータとも大きく違わない。

以上から，発熱量の予測モデルとしては官能基中の結合を変数として用いる事が有効と考えられる。

3.2.2　予測式

官能基中の各結合別の数を中心とした説明変数のモデルで，化学的にもある程度納得できる結果を得る見通しがついたので，ここではそのモデルを使ってベンゼン単環化合物全体（307 件）の解析を行った。

その結果，予測式の良さをあらわす寄与率はR^2=66.9%とかなり小さく，必ずしも良い予測式が得られなかった (Table 9)。

それでも得られた各因子の回帰係数は二置換ベンゼンを対象とした結果と大差ない値が得られ，このモデルを用いた予測式の係数は比較的安定していて，今後データが更に追加されていっても其ほど大きく変化しないと期待される。このことから，未知の化合物の発熱量の予測モデルとしては大きな誤りはないと思われる。

問題は寄与率が小さい事なので，どのような構造の化合物が上手く予測できないかを，残差（実測値－予測値）の分布から調べた。すると，残差の大きな（50kcal/mol 以上）化合物の構造上の特徴として，ベンゼン環に直接結合したハロゲンを含んでいる化合物が非常に多い事がわかる。(Table 10)

そこで，この対策として環に結合しているハロゲンの数を変数として追加して，再度解析を行った。寄

Table 12 Estimate decomposition heat(kcal/mol) from a result of multivariate regression (benzene derivatives).
発熱量の予測値－結合の種類＋環結合ハロゲン
(ベンゼン単環誘導体)

Chemical Substances	Functional Group	Obs.	Pred.
Mono-substituted nitrobenzenes			
p-nitrotoluene	$-CH_3$	54.1	66.7
p-nitrobenzoic acid	$-(C{=}O)OH$	67.8	70.4
p-nitrobenzamide	$-(C{=}O)NH_2$	76.7	66.1
p-nitrophenylhydrazide	$-NHNH_2$	91.1	98.9
parathion	$-OP(OC_2H_5)_2S$	140.0	134.7
Another benzene derivatives			
benzoyl chloride	$-(C{=}O)Cl\ -H\ -H\ -H$	67.8	51.2
3,4-dinitrobenzylalcohol	$-CH_2OH\ -NO_2\ -NO_2\ -H$	164.0	136.5
p-chloroanilidophosphoric acid	$-NHPO(OH)_2\ -Cl\ -H\ -H$	106.0	116.8
3,4,5-trimethoxycinnamic acid	$-CH{=}CHCOOH\ -OCH_3\ -OCH_3\ -OCH_3$	54.0	52.5
2,4,5-trichlorobenzenesulfonyl hydrazide	$-SO_2NHNH_2\ -Cl\ -Cl\ -Cl$	168.0	113.2

Table 13 Typical chemicals with large error.
残差の大きな化合物

Chemical Substance	Functional Group	Obs.	Pred.
5-chloro-2-nitrobenzotrifluoride	$-NO_2\ -Cl\ -CF_3\ -H$	9.5	86.9
2,6-dinitrobenzoic acid	$-NO_2\ -NO_2\ -COOH\ -H$	52.1	136.5
1-chloro-3,4-dinitrobenzene	$-NO_2\ -NO_2\ -Cl\ -H$	81.7	153.0
4-fluoro-3-nitrophenylazide	$-NO_2\ -N_3\ -F\ -H$	218.0	161.4
2-chloro-3,5-dinitrobenzoic acid	$-NO_2\ -NO_2\ -COOH\ -Cl$	245.0	157.3
2,4,6-trinitrotoluene	$-NO_2\ -NO_2\ -NO_2\ -CH_3$	292.0	198.3

与率は$R^2 = 73.5\%$といくらか改善された。その予測式 Table 11 に，また予測値の例は Table 12 に示している。しかし，予測式の良さをあらわす寄与率はまだ小さく，実際に未知化合物の危険性（発熱量）を有効に予測するのはまだ無理がある。

そこで，再び残差（実測値－予測値）の分布を調べ，残差の大きな（絶対値 50kcal/mol 以上）化合物についてその構造上の特徴をみると次の事がわかる。まず予測値の方が小さい化合物（残差が正）の特徴は，複数の塩素原子，あるいは複数のニトロ基を含む事が多い。また予測値の方が大きい化合物（残差が負）をみると，こちらも複数のニトロ基を含んでいる事が多い。そして，複数のニトロ基を含む化合物で，残差の予測値の方が大きい化合物と小さい化合物とをそのニトロ基以外の官能基で比較すると，官能基の種類に大きな違いはなく，異なるのはその組み合わせや結合位置だけである (Table 13, Fig. 10) 。

このことは，今後予測式の寄与率をさらに上げる為には，説明因子に上記の両者（残差が正の化合物と負の化合物）を区別できる因子（上記の官能基の比較からは，官能基間の相互作用を示す因子と考えられる）を追加する必要があることを示している。現在のところ，どのような説明因子を追加すれば良いかはまだ明らかではなく，今後の検討が必要である。

3.3 発熱開始温度の予測

3.3.1 モデルの検討

発熱開始温度について，一置換ニトロベンゼンを対象に，発熱量のモデルとして適していた官能基の種類と官能基中の結合別の数を説明変数としたモデルで解析を行った。

1. 官能基の種類を中心としたモデル
   (Table 14, $R^2 = 71.4\%$)
2. モデル 1 に官能基中の結合別の数を追加したモデル
   (Table 15, R^2=60.3%)

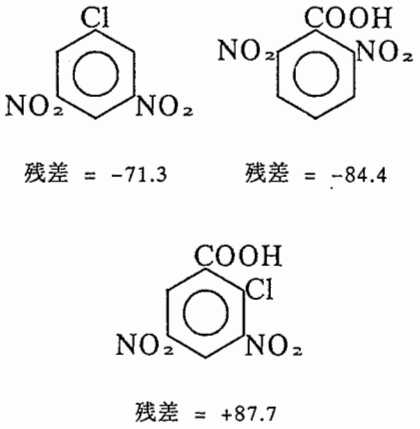

Fig. 10 Typical chemical structure with large error.
残差の大きな化合物の構造式の例

Table 14 A result of multivariate regression analysis for exothermic onset-temperature by kinds of functional group.
(mono-substituted nitrobenzene)
発熱開始温度の予測 1 ー官能基の種類
(一置換ニトロベンゼン)

Variables	Coefficient	Std.Dev.
Const.	291.8	28.8
−(C=O)−(Carbonyl)	+69.6	23.5
−(C=O)NHNH$_2$	−62.6	41.4
−Y(Amine, Ether)	+47.3	25.3
−NN−(Hydrazine)	−118.9	18.5
O−P(Phosphate)	−45.6	54.2
Position(metha,para)	+18.2	27.7

なお，この中で 1 のモデルは，発熱量の解析の場合と同様，該当する官能基の有無を質的変数として，数量化理論第一を利用して回帰分析を行っている。

発熱量を予測した解析結果と違い，発熱開始温度の予測式の寄与率は官能基の結合の種類を説明変数としたモデルよりも官能基の種類を説明変数としたモデルの方が良い。

化合物の発熱量は，分解する官能基の各部分が持つエネルギーの総和で表されると考えられるので説明変数として官能基の結合の種類を用いたモデルが有効であったが，それに対して発熱開始温度は，化合物中に含まれる複数の異なる結合の中で，その結合エネルギーの最も小さい部分の分解が始まる温度と考えられる為，結合を説明変数とするモデルでは上手く説明できないと考えられる。良いモデルの探索は今後の課題である。

4. 判別分析の結果

4.1 発熱量を基にした危険性の判別分析

ここでは，化合物の危険性判定のもうひとつの方法，即ち化合物の構造の情報から直接その危険性を判定する予測式を判別分析を用いて求めることを試みた。現時点では発熱開始温度の適当なモデルがまだ見つかっていないので，今回は発熱量を用いて分類された危険性についてのみ解析をおこなう。ただし最終的には，予測指標として挙げた感度の指標（発熱開始温度）及び威力の指標（発熱量）のそれぞれについて，データの得られている化合物それぞれが危険か否かの判定をして，その両者から Fig. 4 の分類で各化合物の危険性を分類し，それを用いて判別分析を行うべきである。

この判別分析においては，予測モデルとして，回帰分析で用いた官能基中の結合の種類を説明変数としたモデルを使って解析を行う事とする。回帰分析では，発熱量の単位として 1 モル当たりの発熱量を用いたが，現実に化合物の危険性を評価する際には，単位重量当たりの方がより良く現実の危険性を表すと考えられるので，危険性の判定する際の発熱量には単位として cal/g を用いた。

化合物の危険性判定のための基準値の選び方は色々な方法があると思われるが，ここでは ASTM の CHETAH プログラムの化学物質の発熱に係わる危険性分類の数値として 300cal/g と 700cal/g が用いられていることから，この低い方の数値を目安に，CHETAH の値が DSC の実測値より大きい事を勘案して，200cal/g を危険物か否かの判断基準とした。

解析を行う化学物質のグループは，ここでは二置換ベンゼンを対象とした。そして，それぞれの化合物を発熱量 (cal/g) の値から，200cal/g を基準として危険物か否かに分類し，そのデータを官能基中の結合の種類を説明変数としたモデルで分析した。

解析結果は Table 16 に，またその判別結果は Table 17 に示してあるが，正答率としては 90%以上の値が得られた。しかし，その結果について化学的に解釈するのは容易ではなく，モデルの再検討が必要と思われる。

また，特に誤りの中でも危険物が非危険物として判断される誤りは，その逆の場合と比べ問題であるので，非危険物が危険物と判断される誤りは増えて

Table 15　A result of multivariate regression analysis for exothermic onset-temperature by kinds of bonding. (mono-substituted nitrobenzene)
発熱開始温度の予測 2 －結合の種類
(一置換ニトロベンゼン)

Variables	Coefficient	Std.Dev.
Const.	+262.8	40.5
C−C	−16.3	11.5
C−N	+53.4	15.3
N−N	−145.9	19.8
N=O	+34.3	19.4
O−P	−25.0	9.4

も，全ての危険物が危険物と正しく判断されるように何らかの重みづけを行う必要がある。さらに，基準値付近の化合物は発熱量の測定誤差のため，危険物とするか非危険物とするか微妙なものがあり，その判定結果がそのまま判別式の精度と関係する。この点については，例えばファジー理論を応用した解析方法の応用が必要であろう。

5. まとめ

本報告書では，多変量解析の手法を用いて，化学物質の危険性をその分解熱と発熱開始温度から予測する予測式および判定式を求めることを試みた。

回帰分析の結果では，発熱量の予測式として，目的変数に官能基中の各種の結合別の数をとることで，寄与率$R^2 = 73.5\%$(Table 11，対象・ベンゼン単環化合物　307 samples) の予測式が得られた。また，発熱開始温度については，回帰分析による予測式を求めるために，目的変数の検討をしたが良い目的変数が見つかっていない。さらに，化合物の危険性のランクを構造式から直接求める判別式を，発熱量を基準に危険性の判定をしたデータを用いて判別分析で求めた。正答率 90%以上の結果が得られたが (Table 17)，判別係数の化学的な意味づけが難しく，さらに検討が必要である。

今回得られた予測式，判別式では，実際に予測に使用するにはまだ不充分であるが，今後の検討によって，現実の使用に耐えうる予測式が得られる可能性は充分あることが示されたと思う。今後，より良いモデルの検討および分析手法の改良等により，予測手法の確立を図りたい。

Table 16　Discrimination coefficients for the determination of dangerous chemicals based on decomposition heat. (di-substituted benzene)
危険性の判別式
発熱量基準／結合の種類
(二置換ベンゼン)

Variables	Coefficient
Const.	+5.61
C−Cl	−2.91
C−Br	−2.82
N−N	−3.62
N−O	−12.26
NO_2	−10.30
O−O	−12.26
O−P	+10.30
O=P	−32.86
S−O	+3.94
S−P	−32.86
P−Cl	+10.30

Table 17　Classification of dangerous chemicals by discrimination analysis. (di-substituted benzene)
判別式による測定結果
発熱量基準／結合の種類
(二置換ベンゼン)

実測による分類			判定式による分類		
			非危険物	危険物	計
非危険物			56	6	62
危険物			5	61	66
計			61	67	128
正答	117	91.41%			
誤答	11	8.59%			

また，当然のことながら，ここで述べてきた多変量解析による予測式は，これのみで化合物の危険性の事前評価を済ませる事ができるというものでは無く，理論的な方法の計算との併用で行われるべきである。また，それらを総合的に評価して，実際の試験による危険性評価試験を行う時の優先順位を決定すべきである。

(平成 3 年 4 月 22 日受理)

参考文献

1)　藤本康弘・安藤隆之・森崎　繁：DSC データの収集及びその簡単な解析－第 22 回安全工学

研究発表会要旨集，(1989)
2) 藤本康弘・安藤隆之・板垣晴彦・森崎　繁：ベンゼン単環誘導体のDSC測定結果とその化学構造との相関－第23回安全工学研究発表会要旨集，(1990)
3) (財) 未来工学研究所編：化学物質の危険性予測システムの開発研究－昭和60，61年度労働省委託研究報告書 (1987, 1988)
4) 森崎　繁：不安定物質の熱分解危険性に関する研究－東京大学学位論文，(1985)
5) 森崎　繁・安藤隆之：反応性物質のDSCデータ集－産業安全研究所安全資料，RIIS-SD-87, No. 1 (1987)
6) 安藤隆之・藤本康弘・森崎　繁：反応性物質のDSCデータ集 (2) －産業安全研究所安全資料，RIIS-SD-89 (1990)
7) T. Ando, Y. Fujimoto and S. Morisaki: Journal of Hazardous Materials, printing
8) 芳賀敏郎：多変量解析入門，(1989), (財) 日本情報処理開発協会
9) 柳井晴夫・岩坪秀一：複雑さに挑む科学－多変量解析入門，(1976), 講談社
10) 杉山高一：多変量データ解析入門，(1983), (株) 朝倉書店

付　録

本研究で用いた多変量解析の手法の概要

重回帰分析の概要

y(例えば，発熱開始温度）は変数$x_1, x_2, x_3, \cdots, x_p$ (例えば，官能基の数，官能基の種類別の数）の影響を受けて変化するものとする。その関係は一般に式

$$y_i = f(x_{i1}, x_{i2}, x_{i3}, \cdots, x_{ip}) + e_i \qquad (2)$$

で表せる。$x_1, x_2, x_3, x_4, \cdots, x_p$が同じならばいつでも一定の$y$が得られるわけではなく，ばらつきを持っており，それを表すのがe_iの項である。

$\{x_{i1}, x_{i2}, x_{i3}, x_{i4}, \cdots, x_{ip};\ y_i\}(i = 1 \cdots n)$ のデータから，関数fを推定してx からyを予測したい時に重回帰分析が用いられる。この時，$x_j(j = 1 \cdots n)$が「説明変数」，yが「目的変数」である。

誤差e_iについては，次の四つの仮定がおかれている。

a) 不遍性；期待値が零になる。
b) 独立性；e_iと$e_{i'}$(但し$\neq i'$）は互いに独立（無相関），また説明変数も独立。
c) 等分散性；ばらつきはx, yによらない。
d) 正規性；e_iは正規分布に従う。

関数fとして最も標準的なものは

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i \qquad (3)$$

である。データ(x_i, y_i)から，βを推定し，yを予測する式

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_p x_{ip} \qquad (4)$$

を求める。ここに$\hat{y}_i$はy_iの推定値である。$b_0, b_1, \cdots$はβの推定値で，偏回帰係数である。偏回帰係数は，yの実測値と推定値の差の二乗の和

$$\begin{aligned} S_e &= \textstyle\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ &= \textstyle\sum_{i=1}^{n} [y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2} + \\ &\qquad \cdots + b_p x_{ip})]^2 \end{aligned} \qquad (5)$$

を最小にするように決める。S_eが「残差平方和」であり，この方法が「最小二乗法」である。

$y_i, \hat{y}_i, S_e$の関係を Fig. 11 に示す。

ここで，yの総平方和 (S_T, Fig. 12) と残差平方和S_eとの差は，yの変化の内x の影響として説明できる部分で，回帰による平方和S_R，または単に回帰平方和と呼ばれる。すなわち

$$\begin{aligned} &S_T\,(\text{総平方和}) = S_R\,(\text{回帰平方和}) \\ &+ S_e\,(\text{残差平方和}) \end{aligned} \qquad (6)$$

のように，総平方和が二つの成分に分解される。それぞれの成分が総平方和に占める割合はそれぞれ，寄与率，誤差率である。

寄与率は一般にR^2で表される。単回帰式の時は寄与率はx とyの単相関係数rの二乗に等しくなる。

$$R^2 = 1 - \frac{S_e}{S_T} \qquad (7)$$

実際に回帰式を求める際には，説明変数の候補が沢山ある場合，その全部を使って重回帰式を作るのは好ましくなく，できるだけ少数の説明変数を選んで予測力の高い予測式を作る。これが「変数選択」である。この方法としては全ての変数の組み合わせを求めて，その中から良い予測式を見つける方法もあるが計算量が多すぎる。そこで，ここでは変数を増

減しながら予測式を評価して最適な予測式を求める，逐次選択法を利用する。

その場合の変数選択の基準として，ここでは自由度二重調整済寄与率 (R^{**2}) を利用する。これは，

$$R^{**2} = 1 - \left(\frac{S_e}{S_T} \times \frac{n+q+1}{n+1} \times \frac{n-1}{n-q-1}\right) \quad (8)$$

q : 選択された変数の数

で計算される。そして，変数を一つずつ予測式に取り込んでいって，R^{**2}の最大になる予測式を求める。実際にはR^{**2}がほとんど同じ値を持つ予測式がいくつか得られるのでその中から，

a) 対応する専門分野の知識で説明し易い
b) 計測または調査が容易
c) 時系列予測の場合は先行性がある

といったことを評価して最も良い予測式を選ぶ。

判別分析の概要

回帰分析は，説明変数$x_1, x_2, \cdots, x_p$によって，量的変数である目的変数yを予測する式を求める手法であった。予測の対象が質的変数である場合，すなわち，説明変数$x_1, x_2, \cdots, x_p$の知られている対象がどのカテゴリーに属するかを予測する方法が「判別分析」である。

ここでは，予測対象を二群に分ける場合を例に概略を述べる。この場合，どの群に属するか不明の対象を，その説明変数の値が，二つの分布の平均のどちらに近いかで判別する。ここで比較の為の距離は，次の式で表される様にそれぞれの群の標準偏差を単位として表す。なお，ここでは，x の分布が正規分布で二群の分散及び相関係数が等しい事を前提としている。

二つの変数の間に相関が無い場合は，平均を$\tilde{x}_1, \tilde{x}_2$とし，それぞれの軸上の標準偏差をs_1, s_2とすると，ある点 (x_1, x_2) と平均との距離は次のように表される。すなわち，それぞれの軸上の平均との基準化距離u_1, u_2は

$$u_1 = \frac{(x_1 - \tilde{x}_1)}{s_1}, \quad u_2 = \frac{(x_2 - \tilde{x}_2)}{s_2} \quad (9)$$

なので，距離Dの二乗は，$D^2 = u_1^2 + u_2^2$と表される。ここで，D^2をマハラノビスの平方距離と呼ぶ (Fig. 13)。

二つの群の平均とのマハラノビスの平方距離をもとめてその差zを求めれば，

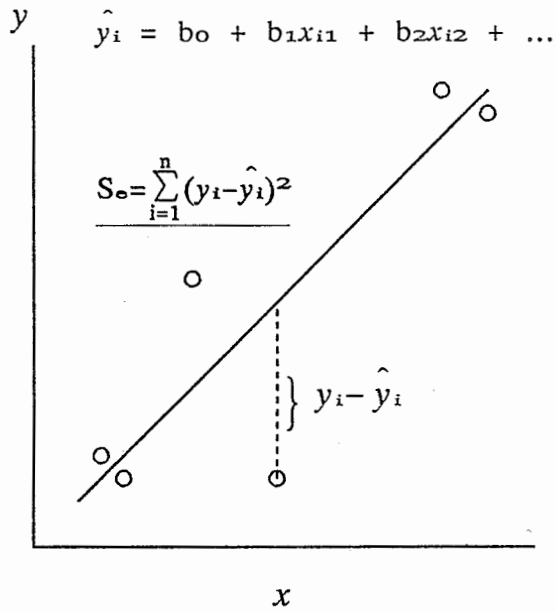

Fig. 11 Relationship between $y_i, \hat{y}_i$, and S_e.
目的変数と回帰分析による予測値および残差平方和との関係

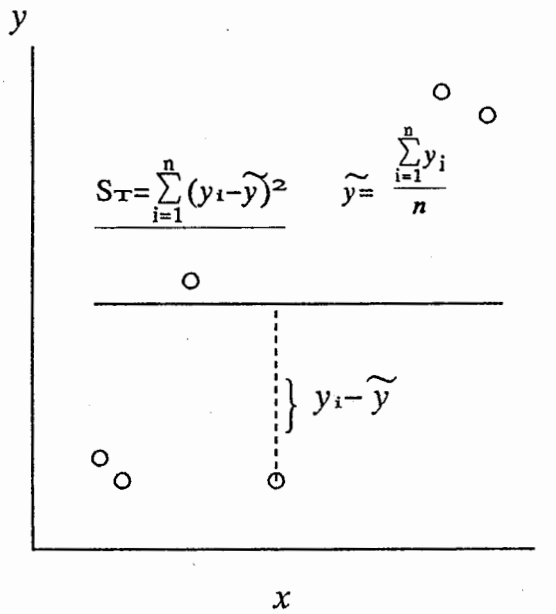

Fig. 12 Relationship between $y_i, \tilde{y}$, and S_T.
目的変数と総平方和との関係

$$z = \frac{(D_2^2 - D_1^2)}{2} \quad (10)$$

zの正負によって，二つの群のどちらに属するかをどちらの群の平均に近いかで判断できる。ここでzが$x_1, x_2, \cdots, x_p$の一次式で表される時，この関数を「線型判別関数」または単に「判別関数」と呼ぶ。

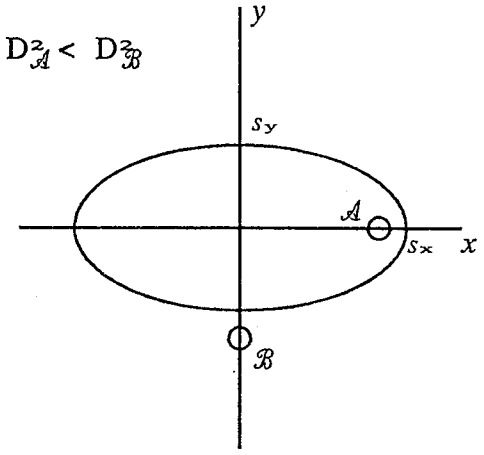

Fig. 13 Mahalanobis' generalized distance D^2.
マハラノビスの平方距離

$$z = a_0 + a_1x_1 + a_2x_2 + \cdots + a_px_p \qquad (11)$$

判別関数を求めるには，重回帰分析の場合と同様に逐次選択法を用いる。その場合の判別関数の良否の基準としては二つの群の平均間の距離D^2をとる。これを判別効率と呼ぶ。また，変数選択の基準は重回帰分析の場合に準じて，自由度二重調整済判別効率 (D^{**2}) という基準値を用いる。

$$D^{**2} = D^2 - \left(\frac{q}{(n-1)} \times \frac{D^2+n(n-2)}{(n_1n_2)} \times \frac{2n}{(n+q+1)}\right) \qquad (12)$$

ここでqは選択された変数の数，またn_1, n_2はそれぞれ二つの群に含まれるケースの数を示す。

最終的な判別式の選択は重回帰分析の場合と同様，固有技術から得られる情報等を考慮して行う。