

## 重みを利用した類似化学反応の検索

4 D - 6

桑原 敏

NTT情報通信処理研究所

## 1. はじめに

化学物質に対する反応経路を既存の反応データから類比的に基づいた類推により求める情報指向型反応設計システムを作成中であるが<sup>[1]</sup>、この方法は、獲得困難な反応理論に基づく理論指向型反応設計システムに較べて、反応事例が化学反応データベースの形で比較的容易に入手できることから、化学設計の分野では比較的有効な方法であると考えられる。

類推を行う場合には、対象の化学物質と既存の化学物質間の類似性を反応の観点から求め、さらに対象の化学物質について反応を適用するの2ステップからなる。

とくに類似性については、化学物質間の化学構造と電気的性質等の部分極大一致による類似度が算定される必要があるが、全ての化学反応についてこれを算定することは、計算量から考えて現実的でなく、既存データをネットワーク<sup>[2]</sup>やフレーム等でカテゴリ分類しておき、これに基づき類似性判定を高速化する方法が採られる。

しかし、このような構造と性質による類似性判定は既存のデータの重要度や信頼度に関係なく類推に使われるために、得られた結果は必ずしも化学者にとって利用価値の高いものとは言えず、システムとしては類似性の高い順に結果を提示し、その中から化学者に選択させる等の手段をとる必要があった。

本研究では、既存のデータに重要度等を示す重みを設け、類似性判定時の類似度算定に利用し、化学者にとって利用価値の高い順に結果が提示できるシステムを作成し、実際に実験したところ、良好な結果が得られた。

以下、本システムにおける類似性判定の概要と、これを用いたシステムの効果について、実験した結果について示す。

## 2. 類似性判定の概要

本システムにおける類似性判定の類似度の計算には、化学反応の1つ1つを、あらかじめ化学物質の構造特徴や、性質特徴を特徴認識した結果をインデックス分類して、化学物質毎のフレームのインスタンスに表現したフレーム形式の知識ベースを用い、さらにインスタンスには反応条件と重みを付加した。(図1)

類似度の算定には、対象とする化学物質の特徴とフレームに記録された特徴間の一致度をスコアリング方式により、

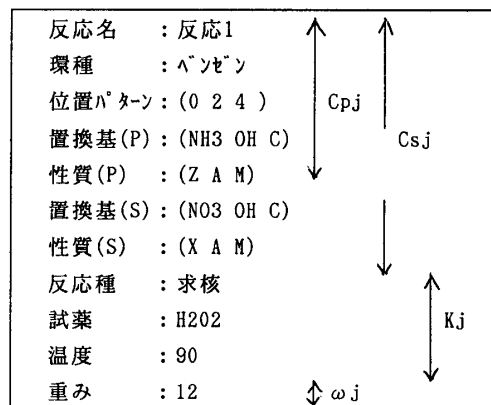


図1 化学反応フレームの例

集計することによって得ることとした。

ここで注意しなければならないのは、化学物質の特徴間の一致度を計算する前に、反応中心となる部分構造を一致させておくことが必要である。これには、対象化学物質内の部分構造が既存の反応の部分構造の変化の一部であるかどうかをチェックする機構を設ける必要があり、これによって反応中心を持ち得ない既存反応は、以後の類似反応の候補には採用しないようにした。

重みを利用しないシステムでは、類似度は各特徴間の一致度と特徴毎に定められた定数の積和を、そのまま類似度  $S_{pj}$  とした。

即ち、既存の化学反応が次のような反応方程式  $R_j$  で与えられているとすると

$$R_j = (C_{sj} \ C_{pj} \ k_j)$$

$$\text{但し、} \ C_{sj} = (C_{sj1} \ C_{sj2} \ \dots \ C_{sji} \ \dots \ C_{sjm})$$

$$C_{pj} = (C_{pj1} \ C_{pj2} \ \dots \ C_{pji} \ \dots \ C_{pjm})$$

$$C_{sji}; \text{ 反応前化学物質の特徴 } i$$

$$C_{pji}; \text{ 反応後化学物質の特徴 } i$$

$$k_j; \text{ 反応条件}$$

$$\text{設計対象の化学物質 } C_t = (C_{t1} \ C_{t2} \ \dots \ C_{ti} \ \dots \ C_{tm})$$

の  $S_{pj}$  は以下によって算定する。

$$S_{pj} = \sum_i \alpha_i \times (C_{ti} \ C_{sji})$$

$$\text{但し、} \ (C_{ti} \ C_{sji}); \text{ 特徴間の一致度}$$

$$\alpha_i; \text{ 特徴 } i \text{ の定数}$$

これに対して、重みを利用する場合は、既存化学反応毎に付与された $\omega_j$ を $S_{pj}$ に掛けたものを類似度 $S_{aj}$ とする。

$$R_j = (C_{sj} \ C_{pj} \ k_j \ \omega_j)$$

但し、 $\omega_j$ ; 既存化学反応の重み(1~100)

とすると、改良された $S_{aj}$ は以下で示される。

$$S_{aj} = \omega_j \sum_i \alpha_i \times (C_{ti} \ C_{sji})$$

類推は、以上の類似度算定結果から得られた類似度の大きい順に、既存化学反応を設計対象の化学物質に適合するように修正することによって得ることができる。

即ち、類推によって得られた設計結果の $R_t$ は次式で示される。

$$R_t = (\text{modified } C_{sj} \ C_{tj} \ k_{j \text{ new}} \ \omega_j)$$

ここで、modified  $C_{sj}$ は $R$ の反応中心を保存しながら $C_{sj}$ を $C_t$ に従い修正したものである。<sup>[1]</sup>

### 3. $\omega$ のチューニングの試み

各反応毎の $\omega_j$ については初期値を反応事例のフレーム構築時に専門家によって設定することとしたが、新たな試みとして、反応事例フレームの利用状況に応じて、 $\omega_j$ を自動的に変化させ、その効果を試験的に見てみることにした。

即ち、実際に設計時において、反応事例フレームの利用回数をシステムが記録しておき、この利用回数に応じて、各事例の $\omega_j$ を増減するものである。具体的には、利用度の低い事例については、 $\omega_j$ の減少を、また利用度の多い事例については、 $\omega_j$ の増加をダイナミックに行い、利用が進むにつれ、類似判定がチューニングされるように図ったものである。

本システムで採用した、 $\omega_j$ の増減量 $\Delta \omega_j$ の計算は以下の通りとした。

$$\Delta \omega_j = 99 / (\max \omega - \min \omega) \times (\omega_j + 1 - \min \omega) + 1 - \omega_j$$

但し、 $\omega$ は全 $\omega_j$ の集合であり、max、及びminはそれぞれ $\omega$ の最大値、最小値を示す。

### 4. 実験結果、及び考察

以下の3条件について実験比較した。

- ① 重み無しの場合
- ② 重み有り、重み変更無し
- ③ 重み有り、重み変更有り

本実験では化学反応を芳香族性単環の置換反応に限定した。また、初期の反応データ数は約300個用意し、実験は既存の代表的反応を20個選択し、各々をフレームから除去し、これが類推後、上位5位以内にある場合に類推が成功したとして成功率を判定した。

また、重み変更については、さらに100個の反応設計

を実施後上記と同様にして、成功率を判定した。各インデックス毎のパラメタ $\alpha$ については①の実験で比較的良い結果を示した数値を用いた。

図2にその結果を示す。これから、①が成功率20%程度であったのに較べて、②では約2倍改善され、また③では約3倍に改善される。追加100個の設計後では①では、ほとんど成功率は改善されなかったが、②では10%程改善された。これにより、重みを付与して類似性判定することにより、類推による反応設計の実用性が格段に向上することが分かった。

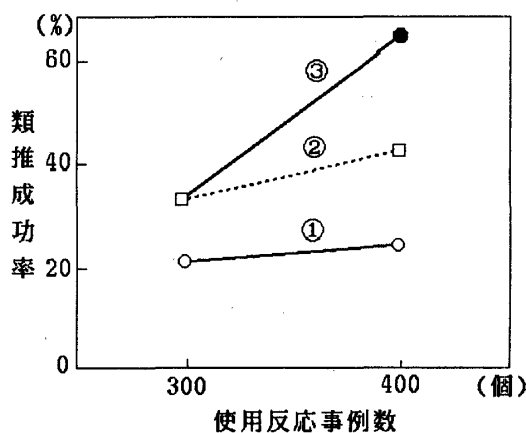


図2 類推実験結果

### 5. まとめ

以上、類似性判定の改善方法を示し、それを適用した類推システムが単純な反応ではあるが、実験により有効であることが分かった。化学の分野では反応事例が記号の形で、データベース化されており、本報告のような類推手法を利用してもかなりの設計支援システムとして利用可能であると考えられる。但し、今後、実用性の観点からは、多様で複雑な反応に対する検証を経る必要があり、そのためには、特徴認識プログラム等における特徴のインデックスの捉え方、かつこれに対応した類似性を決める $\alpha$ の決定に、現状の試行錯誤的な方法でなく、しっかりとした化学ドメインの方法論を検討する必要があると考える。

なお、本研究は、昭和61-昭和63年度科学技術振興調整費による「化学物質設計等支援のための知識ベースシステムに関する研究」の一環として、行われたものである。

### [参考文献]

1. 桑原敏, 山崎毅文, “類推に基づく化学知識獲得方法”, 電子情報通信学会春期全国大会(1989), D-392, 6-112
2. 岡田孝, 河井俊一, “化学構造を対象とする類推システムの開発”, 1989年度人工知能学会全国大会(第3回), 11-12