

高分子の C-13 NMR 化学シフトエキス パートシステム (第 2 報)

巽 久行*・西岡 篤夫**・木澤 誠*

An Expert System of C-13 NMR Chemical Shift of Polymers (2nd Report)

Hisayuki TATSUMI, Atsuo NISHIOKA and Makoto KIZAWA

Abstract

This paper is the second report of a study for developing an expert system of polymers using carbon-13 NMR chemical shift data. The system adopts the inference system of the recognize-act cycle type using frame representations which is an extension of that of the AND/OR tree type using rule representations adopted in the first report. The data of shift values, intensities, shoulders and substructures of carbon at each peak are successively compared with respect to locations, spectral peaks, existence of neighbouring peaks, and the features of spectral patterns with those of unknown polymers which are expected to be identified, and the names of relevant polymers are displayed in the order of certainty. This inference process is divided into four cases according to the degree of sophistication, of which the final one especially gives satisfactory results.

1. はじめに

近年計算機や超伝導コイルの発達に基づく核磁気共鳴 (NMR-Nuclear Magnetic Resonance) 装置の進歩は目覚ましく、その結果 NMR スペクトルの確度は著しく向上して高分子の分野で重要な研究手段として多用されている。特に Carbon-13 (C-13) NMR スペクトルは、高分子化合物の主要な構成要素である炭素原子に着目して分子構造情報を得るもので、最も重要な分析法の一つとなっている。著者の一人は以前から高分子の C-13 NMR 化学シフトデータベースの構築¹⁻³⁾を行っており、化合物名や構造に関する情報から化学シフトデータを検索するシステム^{4,5)}を実現させているが、逆に化学シフトから高分子名を同定させるシステムは今の所著者らの知る限りではほとんど研究されていない。本報告は、この作業を行うエキスパー

トシステムを得ることを目的としたもので、前報告⁶⁾で作成した小規模の AND/OR 木型のプロダクションシステムを、フレーム型知識表現システムに拡張したものである。作成したエキスパートシステムの知識ベース内には、代表的な 68 種類の高分子化合物がフレーム知識として蓄えられており、入力された未知化合物の C-13 NMR 化学シフトデータと比較推論して、確からしい順に候補高分子名が返答される。

このエキスパートシステムは実験的なもので本格的な実用には至らないが、問題点を探ることに重点を置いて簡単なシステムから作成し始め、順次改良を加えながらシステムを構築した。知識表現と推論機能は単純であるにも拘らず概ね満足出来る結果が得られており、C-13 NMR スペクトル以外の棒状スペクトルに対しても応用可能なように設計されている。

2. システムの概要

開発中の「高分子の C-13 NMR エクスパートシス

平成 3 年 9 月 30 日受理

* 情報工学科

** 前情報工学科

テム」は、問題領域として専門知識を多少とも持ったシステムであるが、十分に複雑な問題を解決しているとは言いがたいので完全なエキスパートシステムではない。しかし、それを目指す初期段階のシステムであり、現在 NTT 基礎研究所が開発した汎用性のあるエキスパートシステム構築支援ツール「KBMS」を用いて LISP 言語で記述されたものが、LISP マシン ELIS 8150 上で動作している。以下、システムの全体構成を図 1 に示し、各機能の開発概要を順に列挙する。

(1) 知識ベース

事実に関する知識、判断に関する知識、及び一部推論に関する知識からなる。本システムでは、事実に関する C-13 NMR スペクトル知識をフレーム型知識表現で、推論に利用される判断に関する知識をルール型知識表現で記述している。フレーム型知識表現は、スペクトル情報の枠組みを記述するクラスフレームとその枠組み形式で表現される個々のスペクトルを記述するインスタンスフレームからなる階層型知識である。

ルール型知識表現は、知識をルールの集合で表わす。現システムではルールは全て前向き型で記述しており、関連のあるルールはルールセットでグループ化され、ルールセット内のルールには優先順位が付けられている。後向き型のルールは推論過程説明に利用する予定であり、現在作成中である。推論に関する知識は、ルールセットの選択、及びルールセット内の各ルールの優先順位選択のみからなる。

(2) 推論機構

フレーム型知識の C-13 NMR スペクトル情報とルール型知識の条件部とを照合して、条件が成立すればルール型知識の結論部を実行するという認識-行動サイクル型プロダクションを行ないながら、結論フレームの生成と絞り込みを行なっている。

(3) 知識獲得支援

C-13 NMR スペクトル情報の全体的特徴を、フレーム知識内に自動獲得している。

(4) 推論過程説明

現在作成中であり、HOW 機能、WHY 機能、及び推論過程で知識ベース内に必要な知識がない場合には逆にユーザに質問する機能等を予定している。

(5) 応用プログラム

主に入力されたスペクトル情報と知識ベース内のスペクトル情報との類似度を LISP 言語で記述している。

(6) ユーザインタフェース

使用マシンはスペクトルの画像情報を取り込むことは出来ないで、視覚的な表示は行なわれていない。全て記号処理のインタフェースからなる。

3. 知識フレーム

通常高分子化合物を構成する原子の種類は水素、炭素等が主なものであり、磁気能率を持つこれらの原子核の同位体 ^1H 及び ^{13}C は核磁気共鳴 (NMR) を示す。ある標準化合物の共鳴周波数 (f) に対する試料の共鳴周波数のずれ (df) の相対値を化学シフトと呼び、 df/f によって定義する。 ^1H と ^{13}C の化学シフトはテトラ

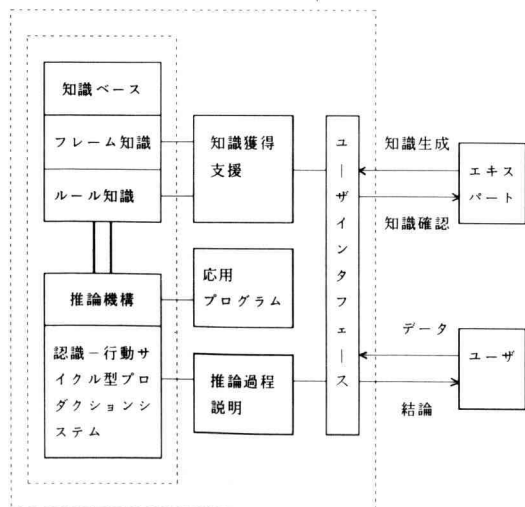
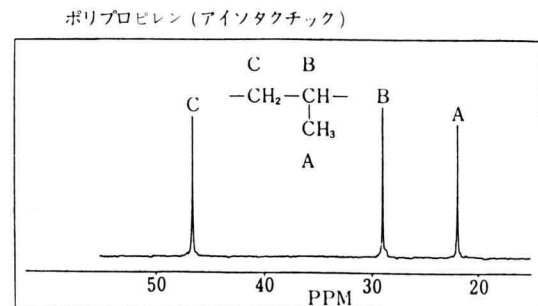


図 1. システム構成図



溶媒: ODCB, 温度: 135°C, 周波数: 50.1 MHz, 積算: 3,400 回
信号 (PPM) A: 21.84, B: 28.83, C: 46.44

図 2. ポリプロピレン（アイソタクチック）の C-13 NMR スペクトル

メチルシラン (TMS) の ^1H と ^{13}C の共鳴周波数を基準として測定されており, ppm 単位で表記されている。この共鳴核のスペクトル位置 (即ち化学シフト値) が隣接する原子との結合状態や分子間相互作用により微妙に変化することから化学シフト値と分子構造との間に関係が生ずる。 ^1H は化学シフト値の存在領域幅が 20 ppm 以内と少ないが, ^{13}C はその幅が ^1H の 10~20 倍 (約 200 ppm) に達しそのスペクトルも尖鋭であるので, C-13 NMR を用いた高分子分析の有用性は極めて高い。図2は代表的な合成高分子の1つであるポリプロピレン (アイソタクチック) の C-13 NMR スペクトル図であり, その分子構造 (CH_2CHCH_3) に対応して右から順 (A~C の順) に CH_3 , CH , CH_2 の各炭素核のシグナルが観測されている。

本研究では, 知識ベース構築に必要な高分子の C-13 NMR スペクトルの原データとして, 測定条件が同じで各炭素核のスペクトル帰属が行われており, かつ代表的な試料が含まれている高分子分析ハンドブック⁷⁾ のスペクトル図を参考にした。これには高分子として代表的な, ポリオレフィン 14 種, ビニールポリマー 15 種, ジェンポリマー及びゴム 10 種, ポリエステル 10 種, 重縮合ポリマー 15 種, その他計 68 種が含まれている。化学シフト値の測定精度は 0.01 ppm 単位であり, 知識ベースとしての信頼性がかなり高いと考えられるので, これをもとにスペクトルのフレーム知識を構成した。

一般に 1 つの高分子化合物の C-13 NMR スペクトルはピークスペクトルの集合 $\{P_s\}$ (但し $1 \leq s \leq n$) から成っている。本システムでは, 化合物のスペクトル情報は, 個々のピークスペクトル P_s が持つ個別的情

報と, 全体のピークスペクトル $\{P_s\}$ が持つ全体的情報に分かれていると考えている。基準値として採用した化合物の原データに関するこれらの情報は, 数値化または記号化されて知識ベース内の知識フレームに格納されている。最初に, ピークスペクトルの個別的情報に関する知識フレームを構成する。

ピークスペクトル P_s の個別的情報には, P_s のシフト値, 強度, 肩及び炭素種等の情報が考えられる。 P_s のシフト値 (shiftval) とは, TMS 基準値を 0 ppm としたときの化学シフト位置であり, 以下これを P_s^{SV} と記す。

P_s の強度 (intensity) とは, ベースラインから最強ピークスペクトルに対する相対強度比とする。即ち P_s の高さを $|P_s|$, 集合 $\{P_s\}$ のうち最も強いピークの高さを $|P_{\max}|$ とすると, P_s の強度は $|P_s|/|P_{\max}|$ であり, 以下これを P_s^{IN} と記す。

尚, この強度 P_s^{IN} は区間に分割 (例えば α 個) して各区間に整数値 $1 \sim \alpha$ を割当てる場合もある。これを, 強度のレベル区間値と呼んでいる。

P_s の肩 (shoulder) とは, 裾に弱いスペクトル分岐が見られる場合で, 原データでは 1 本のピークとして観測されていないが将来単独ピークとして観測される可能性があるものをいう。現システムでは, 右に肩有り $\{r\}$, 左に肩有り $\{l\}$, 左右に肩有り $\{r, l\}$, 及び肩無し $\{\}$ の 4 つの場合で定義されており, 以下これを P_s^{SH} と記す。肩はアナログ的な情報であり, その存在判定は知識ベース構築者の判断で与えられたかなり不確かなものである。

P_s の炭素種 (carbon) とは, 着目炭素種及び化学構造式上で隣接する他のスペクトル情報からなる。これ

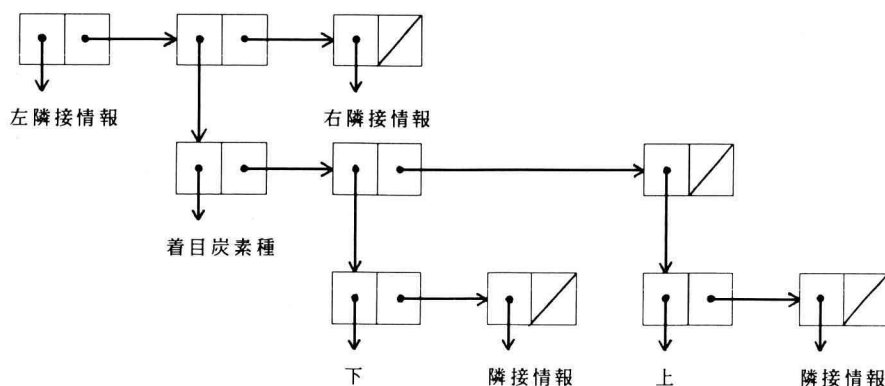


図3. セル構造の例

を LISP のセル構造で表わすと図 3 のようになる。 P_s の炭素種はそのセル構造に対応するリスト

(左隣接情報 (着目炭素種 (上, 下の隣接情報))
右隣接情報)

で, 以下これを P_s^{CB} と記す。

以上がピークスペクトル P_s の個別的情報に関する知識で, 知識集合として

$$\{P_s^{SV}, P_s^{IN}, P_s^{SH}, P_s^{CB}\}$$

で表わされたものが, SHIFT と呼ばれるクラスフレームで定義されている。

例として, 図 2 に示されたポリプロピレン (アイソタクチック) の, ピークスペクトル B の個別的情報は次のようになる。

$$P_s^{SV}=28.83$$

$$P_s^{IN}=1.00 \text{ (レベル区間値は 5, 但し } \alpha=5)$$

$$P_s^{SH}=\{ \}$$

$$P_s^{CB}=(C \text{ (CH (down A)) C})$$

SHIFT クラスには, 68 個の高分子からなる合計 525 個のこのようなインスタンスがスペクトルの個別的情報として与えられている。

次に, ピークスペクトルの全体的情報に関する知識フレームを構成する。ピーク集合 $\{P_s\}$ の全体的情報には, シフト値の存在範囲, ピークの本数及びピークの密集等の情報がある。以降, 便宜的にピーク集合 $\{P_s\}$ を P で表すことにし, 誤りの生じない限り使い易い方を用いることにする。

最初にシフト値の存在範囲 (bounds) とは, 集合 P のシフト最小値を $\text{low}(P)$, シフト最大値を $\text{high}(P)$ で表わすと, $(\text{low}(P), \text{high}(P))$ となるリストであり, 以下これを $\text{bounds}(P)$ と記す。

次にピークの本数 (number) とは, 集合 P の個数である。尚, 強度がレベル区間値で与えられている場合には, $\{\text{区間値 1 の本数}, \dots, \text{区間値 } \alpha \text{ の本数}\}$ となるリストであり, 以下これを $\text{number}(P)$ と記す。

最後にピークの密集 (crowd) とは, 狭い範囲内に隣接ピークスペクトルが存在する場合に定義される。密集情報は, 密集をなすピーク集合の存在範囲とそのピークの本数で与えられる。例えば密集ピーク集合が m 箇所あり, 各密集ピーク集合が C_k (但し $1 \leq k \leq m$ で, $C_k \subset P$) とすると, 集合 P の密集は,

$$((\text{bounds}(C_1), \text{number}(C_1)), \dots, (\text{bounds}(C_m),$$

$$\text{number}(C_m)))$$

となるリストであり, 以下これを $\text{crowd}(P)$ と記す。

密集情報は, 個別的情報である肩の情報と密接な関係にある。肩が単独ピークとして観測されれば, 密集を形成する可能性が高い。

以上がピークスペクトルの全体的情報に関する知識で, 知識集合として

$$\{\text{bounds}(P), \text{number}(P), \text{crowd}(P)\}$$

で表わされたものが, POLYMER と呼ばれるクラスフレームで定義されている。

例として, 図 2 に示されたポリプロピレン (アイソタクチック) の全体的情報は次のようになる。

$$\text{bounds}(P)=(21.84 \ 46.44)$$

$$\text{number}(P)=3 \text{ (レベル区間は } \{-1, -1, -1, -1, 3\})$$

$$\text{crowd}(P)=(\quad)$$

POLYMER クラスには, 68 個の高分子からなるこのようなインスタンスがスペクトルの全体的情報として与えられている。

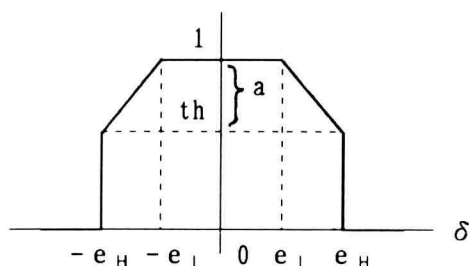
4. 推 論

本システムで使用している推論は主に記号推論, それもかなりパターン・マッチ的な色彩が強い。以下に本研究の推論の方法を述べる。

4.1 基本方針

入力された高分子化合物データのピークスペクトル集合を $\{I_j\}$, 知識ベース内のある高分子化合物におけるピークスペクトルの集合を $\{P_s\}$ とする。以降, 便宜的にピーク集合 $\{I_j\}$ 及び $\{P_s\}$ を各々 I 及び P で表わす。システムは, 入力スペクトル I_j の個別的知識の集合 $\{I_j^{SV}, I_j^{IN}, I_j^{SH}\}$ が与えられると, $\{I_j\}$ の全体的知識の集合 $\{\text{bounds}(I), \text{number}(I), \text{crowd}(I)\}$ を生成して推論に入る。

最初に, 入力された I_j の個別的知識の集合 $\{I_j^{SV}, I_j^{IN}, I_j^{SH}\}$ と知識ベース内にある P_s の個別的知識の集合 $\{P_s^{SV}, P_s^{IN}, P_s^{SH}\}$ との間の, 類似度の強弱を表わす評価を定義する必要がある。 I_j に対する P_s の類似度を $(P_s | I_j)$ で表わすことにすると, この個別類似度はシフト値, 強度及び肩の 3 情報に対して, 各々次のように定義される。

図4. 類似度 $(P_S | I_J)^{SV}$; P_S^{SV} と I_J^{SV} の対応

[シフト値の個別類似度]

シフト値は化合物の測定条件による溶媒効果がある(シフト値が左右にずれる)ので, 台形近似で溶媒効果を考慮した類似度を次のように定義する。

$$(P_S | I_J)^{SV} = \begin{cases} 1 & (|e| \leq e_L \text{ のとき}) \\ a \in (th, 1) & (e_L < |e| \leq e_H \text{ のとき}) \\ 0 & (e_H < |e| \text{ のとき}) \end{cases}$$

但し, $e = P_S^{SV} - I_J^{SV}$ で, 溶媒効果 e_L 及び e_H は $0 < e_L < e_H$ とする。溶媒効果は測定条件により異なるが, 大きくても 1 ppm 未満の値であり, また th (但し $0 < th < 1$) は溶媒効果に基づく類似度のしきい値を表わしている。図4にこの類似度を示す。

[強度の個別類似度]

シフト値の個別類似度が0でない場合に, 次のように定義する。

$$(P_S | I_J)^{IN} = \begin{cases} 1 & (I_J^{IN} = P_S^{IN} \text{ のとき}) \\ b \in (0, 1) & (I_J^{IN} \neq P_S^{IN} \text{ のとき}) \\ 0 & (I_J^{IN} \text{ が unknown のとき}) \end{cases}$$

但し, $b = \min \{P_S^{IN}, I_J^{IN}\} / \max \{P_S^{IN}, I_J^{IN}\}$ とする。

[肩の個別類似度]

シフト値の個別類似度が0でない場合に, 次のように定義する。

$$(P_S | I_J)^{SH} = \begin{cases} 1 & (I_J^{SH} = P_S^{SH} \text{ のとき}) \\ c_i \in (1/2, 1) & (I_J^{SH} \subsetneq P_S^{SH} \text{ のとき}) \\ 1/2 & (I_J^{SH} \text{ が unknown のとき}) \\ d_i \in (0, 1/2) & (I_J^{SH} \supsetneq P_S^{SH} \text{ のとき}) \\ 0 & (I_J^{SH} \cap P_S^{SH} = \phi \text{ のとき}) \end{cases}$$

但し, c_i 及び d_i は, 集合 I_J^{SH} と P_S^{SH} が等しいほど大きな値となる。図5にこの類似度を示すが, 図中で $0 < d_1 < d_2 < 1/2 < c_1 < c_2 < 1$ としている。

以上がスペクトルの個別的知識の3要素 SV , IN

$\begin{matrix} P_S^{SH} \\ I_J^{SH} \end{matrix}$	$\{ \quad \}$	$\{ \text{左} \}$	$\{ \text{右} \}$	$\{ \text{左}, \text{右} \}$
$\{ \quad \}$	1	c_2	c_2	c_1
$\{ \text{左} \}$	d_2	1	0	c_2
$\{ \text{右} \}$	d_2	0	1	c_2
$\{ \text{左}, \text{右} \}$	d_1	d_2	d_2	1

図5. 類似度 $(P_S | I_J)^{SH}$; P_S^{SH} と I_J^{SH} の対応

及び SH に対する類似度の定義であり, これを用いて I_J に対する P_S の個別類似度 $(P_S | I_J)$ は次のように定義される。

[個別類似度]

$$(P_S | I_J) = \sum \{v^* \times (P_S | I_J)^*\}$$

但し, $*$ は3要素 SV , IN 及び SH のいずれかであり, $v^* (\in [0, 1])$ は各要素に付けられた評価係数, \sum は3要素に対する全ての加算を意味する。

更に, 入力情報集合 I に対する知識情報集合 P の全体類似度を $(P | I)$ で表わすことにすると, これは個別類似度から次のように定義される。

[全体類似度]

$$\begin{aligned} (P | I)^* &= \sum_J (\sum_S (P_S | I_J)^*) \\ (P | I) &= \sum_J (\sum_S (P_S | I_J)) \end{aligned}$$

但し, 内側の \sum_S 計算はある入力情報 I_J に対する知識情報の個別類似度を知識情報集合全てについて加算することを, 外側の \sum_J 計算はこの計算を全ての入力情報集合について加算することを表わしている。

上述した類似度の定義により, 入力された各ピークスペクトル I_J が知識ベース内にあるピークスペクトル P_S に類似している程, その個別類似度 $(P_S | I_J)$ は高い値を示す。更に, 入力されたピークスペクトル集合 I が知識ベース内にあるピークスペクトル集合 P に類似している程, その全体類似度 $(P | I)$ は高い値を示す。

ここで仮に入力情報集合 I と知識情報集合 P が一致した場合, 個別類似度は各ピーク毎に理想最高値を示し, 全体類似度も理想最高値を示す。よってシステムに, これらの個別類似度の最高値 $(P_S | P_S)$ 及び全体類似度の最高値 $(P | P)$ の値を前もって知識ベース内に蓄えておけば, 知識ベース内にあるピークスペクトルの候補判定及びピークスペクトル集合の候補判定を行うことが出来る。これは, 次のような一致率 (Consistency) を定義すれば簡単に行える。

[一致率の定義]

知識情報 P_S と入力情報 I_I との一致率 $C(P_S|I_I)$ は、次の式で与えられる。

$$C(P_S|I_I) = (P_S|I_I) / (P_S|P_S)$$

特に $P_S = I_I$ のとき、 $C(P_S|I_I) = 1$

同様に、知識情報集合 \mathbf{P} と入力情報集合 \mathbf{I} の一致率 $C(\mathbf{P}|\mathbf{I})$ は、次の式で与えられる。

$$C(\mathbf{P}|\mathbf{I}) = (\mathbf{P}|\mathbf{I}) / (\mathbf{P}|\mathbf{P})$$

特に $\mathbf{P} = \mathbf{I}$ のとき、 $C(\mathbf{P}|\mathbf{I}) = 1$

また、 $C(P_S|I_I)^*$ 及び $C(\mathbf{P}|\mathbf{I})^*$ も同様に定義される。但し $*$ は個別的情報の 3 要素をなす SV , IN , 及び SH のいずれかである。

一致率の定義から分かるように、この値は類似度計算が正しく類似具合いを保証してくれるときにのみ有効である。これより推論は、一致率が限りなく 1 に近いものを候補スペクトルさらには候補高分子として選択すればよい。

以上、一致率を考慮した候補判定は知識ベース内の知識が理想値であり、推論はその理想値に近づくものを候補として選択する。しかしながら、入力スペクトル情報集合が必ずしも知識ベース内のスペクトル知識情報集合より確度が悪いとは限らない。より高確度の C-13 NMR 測定装置で求められたスペクトルに対しては、理想値と外れてしまうためにシステムは誤った推論を起こす可能性がある。

4.2 モデル別推論

作成したシステムは、その知識ベースの内容と推論方法により段階的に 4 つのモデルに分けられる。モデル 1 及びモデル 2 は知識ベースとして SHIFT クラスに限定したもので、スペクトルの個別の特徴のみで候補高分子を推論する。モデル 3 及びモデル 4 は知識ベースに POLYMER クラスを加えたもので、スペクトルの全体的特徴も考慮に入れて候補高分子の推論を行う。特にモデル 4 は候補高分子の推論評価を行って、結果の妥当性を検証している。以下、その概要を列挙する。

(1) モデル 1

知識情報集合はシフト値集合 $\{P_S^{SV}\}$ のみからなる SHIFT クラスで、推論は $\{P_S^{SV}\}$ に設定された許容誤差範囲内で候補高分子を検索して論理積のみで絞り込みを行っている。このモデルは、入力されたシフト値

集合 $\{I_I^{SV}\}$ に対する誤差範囲内での候補リストの情報と、最終候補を絞りきるまでの所用時間の測定実験に用いた。入力された各シフト値に対して 525 個の全てのインスタンスを検索するので、当然ながら所用時間は入力されたシフト値の数に比例する。許容誤差範囲が ± 0.5 ppm で、シフト値の数が最も少ない高分子 (2 個: シス-1, 4-ポリブタジエン) で約 15 秒、最も多い高分子 (15 個: ポリアミノビスマレイミド) で約 82 秒を要した。許容誤差範囲により候補高分子の数が増大し、ある範囲以上では結果そのものが無意味となる。

(2) モデル 2

知識情報集合はシフト値、強度及び肩の集合 $\{P_S^{SV}, P_S^{IN}, P_S^{SH}\}$ からなる SHIFT クラスで、推論は個別類似度を加算してその大小順に候補高分子を絞り込む。入力情報集合 $\{I_I^{SV}, I_I^{IN}, I_I^{SH}\}$ の中で、強度のレベル区間値が判別し難い場合は区間値の集合あるいは unknown も許される。肩は単にその有無のみを与える。シフト値の許容誤差範囲が $\pm e_H$ (例えば、 $(e_L, e_H) = (0.2 \text{ ppm}, 0.5 \text{ ppm})$, $th = 1 - (e_H - e_L)$) で個別類似度を求め、全体類似度がその理想値の th 倍 (但し、 $0 < th < 1$) 以上あるものを大小順に候補高分子として推論する。誤差範囲が $\pm e_L$ 以内で強度と肩のデータが一致すれば全体類似度は理想値となる。実行結果は、入力シフトの数が多いたときは概ねよい結果を得るが、少ないときは類似化合物との判別が困難となる。

(3) モデル 3

知識情報集合はスペクトルの個別の情報以外に、全体的情報であるシフト値の存在範囲、ピークの本数及び密集の集合 $\{\text{bounds}(\mathbf{P}), \text{number}(\mathbf{P}), \text{crowd}(\mathbf{P})\}$ からなる POLYMER クラスも使用する。密集は $\pm e_H$ 範囲内の低密集及び $\pm e_L$ 範囲内の高密集で定義される。推論はモデル 2 で絞られた候補高分子名を中間仮説とし、入力集合と中間仮説集合の整合性を、入力シフト値の存在範囲、ピークの本数、及び密集の順に調べる。推論は、判定が全て終了しなくても途中で候補高分子が 1 つに絞られた時点で終了する。この方法で、モデル 2 では候補高分子を 1 つに絞り切れなかった未知化合物 (例えばポリプロピレンの例では、アイソタクチック、エチレンとのランダム共重合体及びエチレンとのブロック共重合体の 3 つ) も、判別出来るようになった。所用時間はモデル 2 で 1 つに絞り切れた場合と大差はなく、平均数十秒である。

(4) モデル 4

知識情報集合は、モデル 3 に着目炭素種の個別的情

報 $\{P_S^{CB}\}$ を加えている。この炭素種の情報は、入力情報と知識情報との間の矛盾検出に利用されている。システムは全体類似度 $(P|I)$ が 0 でない場合に新しい POLYMER クラス PC_0 、及び新しい SHIFT クラス SC_0 を生成して類似度の情報を蓄えておく。新しく生成されたクラス PC_0 に入る候補高分子は中間仮説であり、次にシステムはこの中間仮説をもとにして一致率 $C(P|I)$ の計算に移る。一致率が 1 に等しい場合、1 より小さい場合、1 より大きい場合に分類して各々新しい POLYMER クラス PC_1 、 PC_2 及び PC_3 を生成して各クラスの評価に移る。ここで新しく生成された各クラスの評価概要は次のようになっている。

PC_1 クラスは候補高分子の可能性がかなり高い。システムはこのクラスに含まれる高分子が 1 個の場合、他のクラスに一致率の値が 1 に近い高分子が無ければほぼ無条件に候補高分子と推論する。またこの場合、知識ベースの基準値よりも高確度な入力高分子である可能性もあるが、システムにはそれが高確度データであるか否かの判定能力がない。

PC_2 クラスは一致率が大きい程、候補高分子の可能性が高い。システムは全体的情報と合わせて総合判断を行うが、特に強いピークスペクトルの類似度が大きい程、候補高分子と推論する。然しながら、弱いピークでも意味のある場合には問題となる。

PC_3 クラスは入力高分子に、知識ベースにはない新しい密集が出来た場合であり、高確度なデータを持つ高分子である可能性が高い。一致率が 1 をはるかに越える候補高分子は現われないと考えられるのでこの候補判定は難しく、現システムの大きな問題点となっている。

また、一致率が 1 もしくは 1 に近い候補高分子が複数ある場合には混合物高分子の可能性がある。この場合も高確度高分子と同様に複数のクラスに分類されるので、統一的な判定方法は確立されていない。システムは知識ベースの基準値データをもとに推論を行っているので、基準値データを包含するような入力情報に対しては推論アルゴリズムの再検討が必要である。現在、Dempster-Shafer の確率論⁸⁾ 及び fuzzy 推論⁹⁾ でこの問題の解決を試みているが、逆に仮説の数が増大して組合せの爆発が起きており、その計算量から実時間内での推論は不可能となっている。

5. 実行例

モデル 4 を用いて、ポリプロピレン (アイソタクチック: 略記号 pp-isot) の知識ベース内の基準データを、+0.3 ppm ずらしたものを入力データに与えた場合の実行例を図 6 に示す。システム内での溶媒効果を ± 0.5 ppm (但し、 $e_L = 0.2$ ppm, $e_H = 0.5$ ppm) に設定した場合、実行結果は一致率 $C(P|I)$ が 97.999(%) という高い値を示していることがわかる。pp-isot の入力スペクトルは、全てのスペクトルに類似するものがエチレンとのランダム共重合体 (略記号: cep-ran) 及びエチレンとのブロック共重合体 (略記号: cep-

shift検索開始!!

```
1
  22.14 4 3
  ((cep-blk cep-blk/1 1.45
  (pp-isot pp-isot/1 1.45
  (cep-ran cep-ran/2 1.3))
2
  29.13 4 1
  ((pp-isot pp-isot/2 1.45)
  (cep-blk cep-blk/3 1.3)
  (cep-ran cep-ran/4 1.255)
  (cehxe cehxe/4 1.23))
3
  46.74 4 3
  ((cep-blk cep-blk/A 1.45)
  (pp-isot pp-isot/3 1.45)
  (pp-synd pp-synd/4 1.0)
  (cep-ran cep-ran/8 0.91506))
```

shift検索終了!!

```
(pp-isot 4.3499)
(cep-blk 4.1999)
(cep-ran 3.47)
(cehxe 1.23)
(pp-synd 1.0)
```

polymer検索開始!!

```
>>pp-isot
>>cep-blk
>>cep-ran
>>cehxe
>>pp-synd
```

polymer検索終了!!

```
(* リア nt ヲ (7 イソタクチック)
pp-isot (P|I)=7.3499 C(P|I)=97.999)
(* リア nt ヲ (エチレンとのランダム共重合体)
cep-ran (P|I)=6.47 C(P|I)=40.438)
(* リア nt ヲ (エチレンとのブロック共重合体)
cep-blk (P|I)=7.1999 C(P|I)=36.363)
(エチレン-ヘキセン-1 共重合体
cehxe (P|I)=4.23 C(P|I)=19.859)
(* リア nt ヲ (ソジ オタクチック)
pp-synd (P|I)=2.0 C(P|I)=19.802)
```

図 6. モデル 4 を用いた実験結果

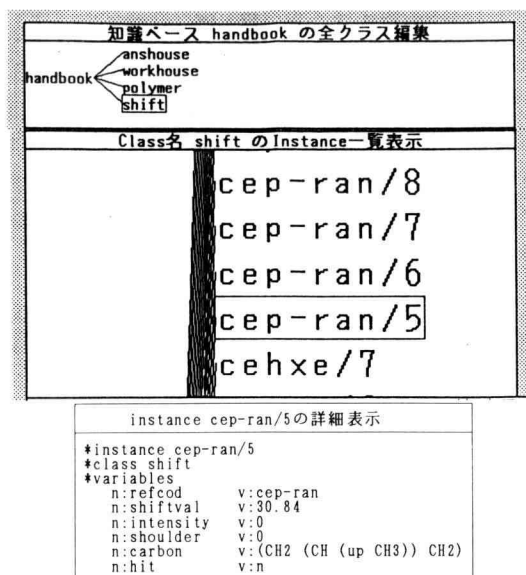


図7. SHIFTS クラスのインスタンスフレームの例

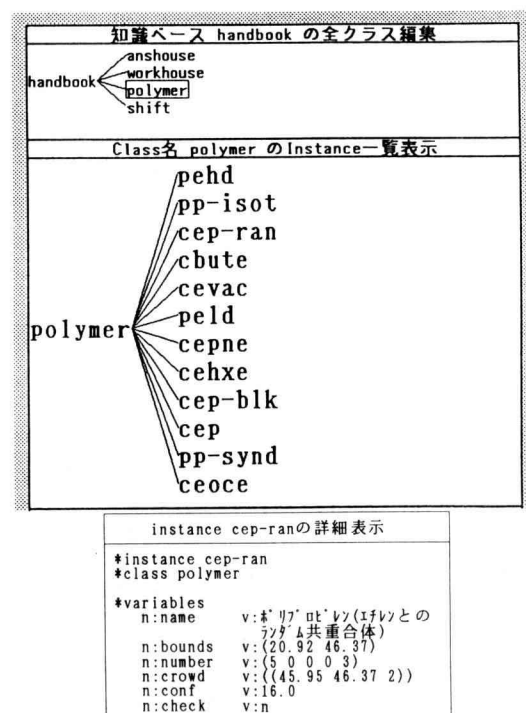


図8. POLYMER クラスのインスタンスフレームの例

表1. 類似度 ($P|I$) と一致率 $C(P|I)$ の比較

	pp-isot		cep-ran		cep-blk	
	$(P I)$	$C(P I)$	$(P I)$	$C(P I)$	$(P I)$	$C(P I)$
実験1	7.50	1.000	6.70	0.419	7.35	0.371
実験2	7.35	0.980	6.47	0.404	7.20	0.364
実験3	6.35	0.847	5.55	0.347	6.20	0.313
実験4	7.35	0.980	6.85	0.428	7.50	0.379
実験5	7.50	1.000	8.40	0.525	9.05	0.457

blk)に含まれているので、全体類似度 ($P|I$) の値はかなり近くなっていることが分かる。また図7及び図8に、実際にシステムで構築した知識ベース内のSHIFTクラス及びPOLYMERクラスの知識フレームを、ポリオレフィン系の12種類の高分子に限定して作成した例を示す。

次に、知識ベース内に含まれるポリプロピレン（アイソタクチック）のデータをもとにした実験結果を示しておく。実験項目は次の5つである。

- (1) 無操作
- (2) 全ての化学シフト値を0.3 ppm ずらした場合（図6の場合）
- (3) 第1ピークスpekトルの強度を1レベル下げた場合（但し、強度を5つのレベル区間値に分けている）
- (4) 第2ピークスpekトルの肩を変化させた場合
- (5) ピークスpekトルを1本増やした場合（但し、強度は最小レベル1に含まれる弱い信号とした）

以上の実験項目別に、図6で候補高分子に挙げたエチレンとのランダム共重合体及びエチレンとのブロック共重合体に対する全体類似度 ($P|I$) と一致率 $C(P|I)$ との比較を、表1に示す。特に実験項目(5)は弱いスペクトルが少し増えても一致率は変化しないという高精度データに対する実験である。然しながら強いスペクトルではこの限りではない。知識ベース内に今まで肩で捉えられていた信号が入力データとして数多く与えられた場合の一致率はかなり減少される。

6. おわりに

未知化合物のC-13 NMR 化学シフト情報から、その化合物名を同定させる機能を持つエキスパートシステムを作成した。本システムは実験的な小規模システムではあるが、知識構造の単純さにも拘らず、今の所

概ね満足出来る推論が可能となっている。また実行時間も数十秒であり、実用時間内で動作している。然しながら本文中で指摘したように問題点も多く見られるので、以下にまとめて列挙する。

(1) 拡張性のある知識構造ではあるが、高分子化学者が描く高分子の知識構造ではない。化合物の分類は一般に複雑であり、階層的でしかも網的でもあるが、本システムは1階層の知識構造でしか表現されていない。立体規則性、共重合、異種結合、混合物その他の要因をどのように扱うのかは今後の課題である。

(2) 知識ベースの基準値の選択を一種類の原データに依存することは、汎用性の低下を招く。測定条件が異なる異種データ¹⁰⁾を積極的に取り入れなければならない。

(3) 現システム内には、エキスパートの判断知識を余り取り込んでいない。このため、スペクトルのパターンマッチだけでは解決出来ない場合が構築中数多く見られた。また化学シフトの加減則等を考慮するなど、深い知識を推論に利用する必要がある。

(4) 知識ベース内にある基準値よりも高精度で測定されたデータに対しては、推論の完全性が保障されていない。更に混合物のデータに対しても、推論が一部不完全となる可能性がある。正確な混合比などを返答するシステムは今後の課題である。

謝 辞

日頃御指導頂く情報工学科長石井博章教授に深謝致します。更にプログラム作成に御協力頂いた平成元年度卒業研究生の須賀正、宍戸正明両君に、及び平成2年度卒業研究生の藤原信一、田中修二両君に深謝致します。

参 考 文 献

- 1) 西岡: 高分子の CARBON-13 核磁気共鳴データベース CNMRP, 京都大学大型計算機センター広報, Vol. 16, No. 5, pp. 253-265, (1983).
- 2) 西岡, 畑田, 藤原, 末広: 高分子の CARBON-13 NMR リレーションアルデータベース SHIFT, 京都大学大型計算機センター広報, Vol. 19, No. 2, pp. 54-66, (1986).
- 3) 西岡, 藤原: 高分子の CARBON-13 NMR スペクトル図形のデータベース化とその検索, 京都大学大型計算機センター広報, Vol. 19, No. 4, pp. 269-275, (1986).
- 4) 西岡, 畑田, 藤原: 文献, 数値及び図形の3種の情報を利用する高分子の ^{13}C -核磁気共鳴データベースシステム, 分析化学, Vol. 36, No. 1, pp. 64-66, (1987).
- 5) 西岡: データベースを含む高分子の CARBON-13 NMR 知的情報処理システム, 平成元年度文部省科研費研究成果報告書(研究課題番号 62550659), (1990).
- 6) 巽, 西岡: 高分子の C-13 NMR 化学シフトエキスパートシステム, 神奈川工科大学研究報告, B-14, pp. 189-197, (1990).
- 7) 日本分析化学会編: 高分子分析ハンドブック, 朝倉書店, (1985).
- 8) 淵, 古川, 溝口: 知識プログラミング, 共立出版, (1988).
- 9) E.H. Mamdani and B.R. Gaines: Fuzzy Reasoning and its Applications, Academic Press, (1981).
- 10) クラレ中央研究所編(植月監修): POLYSPECS-1, —CD-ROMによる高分子NMRスペクトルデータシステム—, 日本電子社, (1989).
- 11) 上野: 知識工学入門, オーム社, (1989).
- 12) 巽, 須賀, 宍戸, 西岡, 木澤: 高分子の CARBON-13 NMR 化学シフトエキスパートシステム, 第4回人工知能学会全国大会, 16-5, pp. 491-494, (1990).
- 13) B.G. Buchanan and E.A. Feigenbaum: Dendral and Meta-Dendral; Their Applications Dimension, Artificial Intelligence, Vol. 11, pp. 5-24, (1978).
- 14) B.G. Buchanan, G. Sutherland and E.A. Feigenbaum: HEURISTIC DENDRAL; a Program for Generating Explanatory Hypotheses in Organic Chemistry, Machine Intelligence, Vol. 4, 12, pp. 209-254, (1969).
- 15) E.A. Feigenbaum, B.G. Buchanan and J. Lederberg: On Generality and Problem Solving; A Case Study Using the DENDRAL Program, Machine Intelligence, Vol. 6, 12, pp. 165-190, (1971).