例外事例を含むＤＢからの知識自動抽出

6C-6

山崎　毅文　桑原　敏　服部　文夫
ＮＴＴ　情報通信処理研究所

１．はじめに

知識システムの高度化、及び大規模化に伴い、要求される知識の量は、拡大傾向にあり、専門家から獲得できる知識だけでは、充分な知識を確保することは困難になりつつある。知識を大量に獲得するためには、知識源として、事例データベース（ＤＢ）を用い、自動的に知識獲得を行なう方法が考えられるが[1]、ＤＢ中に存在する例外事例等の特殊事例の混在が、正しい知識の獲得を阻害するため、これら特殊事例の検出、及び削除が必要である。

本稿では、まず一般的な知識抽出モデルに基づき、ＴＭＳ(Truth Maintenance System)を利用した、例外事例の除去方法について提案する。

さらに、知識抽出システムの具体例として、化学反応ＤＢから置換基の位置選択性知識を抽出した結果について述べ、本提案が有効であることを示す。

２．ＤＢからの知識抽出における問題

２．１　知識抽出モデル

ＤＢからの知識抽出では、ＤＢ内の個々の事例から、考えられる複数の解釈を生成し、それらの解釈のうち、整合性のある組合せを、知識として蓄積する方法が考えられる。

生成された各々の解釈は、幾つかの仮説から構成されており、全ての解釈の組合せについて、仮説間に存在する制約条件が制限になり、その整合性が計算される。

事例の集合を$\mathbb{D}$、それから生まれる解釈の集合を$\mathbb{I}$、解釈Ｉを構成する仮説の集合を$\mathbb{H}$、また仮説間に存在する制約の集合を$\mathbb{C}$とした場合、以下にように表わされる。

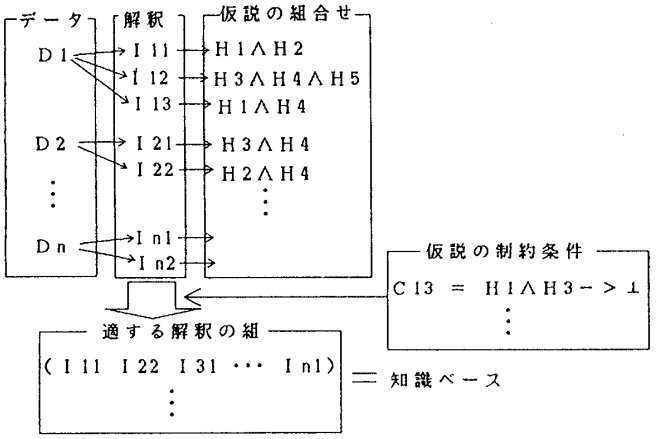
$\mathbb{D} = \{D_i \mid i=1,..,n\}$

$\mathbb{I} = \{I_{ij} \mid i=1,..n\ ,\ j$ ： Diから生成される解釈数$\}$

$\mathbb{H} = \{H_i \mid i=1,..,m\}$

$\mathbb{C} = \{C_{kl} \mid C_{kl} = H_k \wedge H_l \rightarrow \perp\}$

但し、⊥は、論理的矛盾を表わす。



（図１．知識抽出モデル）

図１に示すように、データＤ1から、対象分野の知識に基づいて、３つの解釈Ｉ11,Ｉ12,Ｉ13が生成される。各解釈は、Ｉ11=Ｈ1∧Ｈ2,Ｉ12=Ｈ3∧Ｈ4∧Ｈ5,Ｉ13=Ｈ1∧Ｈ4 のように、仮説集合の要素の連言で構成される。Ｄ1,Ｄ2,...等の各データから生成される解釈の中から一つ選んで、Ｉ11∧Ｉ21,Ｉ12∧Ｉ21等の解釈の組み合わせを生成し、Ｃ13 = Ｈ1∧Ｈ3 -〉⊥ 等の仮説間の制約条件を満たすものを、適する解釈の組として蓄積する。それらの組合せの集合が知識ベースである。

２．２　例外データの存在による問題とその解決策

ＤＢ中に、例外データ、あるいは誤データ等の特殊事例が存在した場合に、前項のモデルにおいて、整合性のある解釈の組合せが生成できないという問題が生じる可能性がある。即ち、解釈のどの組合せを選んでも、制約条件を満たす組合せが存在しない場合である。この解決策として、整合性のある解釈の組合せを阻止するデータを例外データとして、知識抽出の対象から除去すれば、整合性のある組合せを得ることができる。

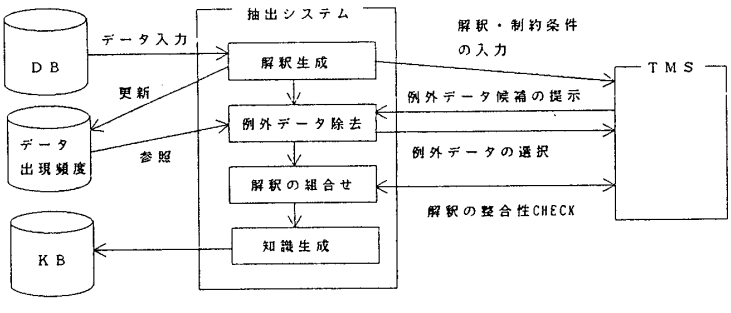
２．３　事例頻度に基づく例外データの除去

整合性のある組合せを阻止する例外データの除去は、まず、解釈と制約条件を用いた論理的整合性の計算により矛盾データの組合せを求めるステップと、次にそれらの矛盾データのうち、ＤＢにおける出現頻度の低いデータを例外データと定め、知識抽出の対象から、除去するステップより構成される。

３．ＴＭＳを用いたＤＢからの知識抽出システム

前章で述べたように、ＤＢから知識抽出を行なう場合には、例外データの除去を行う必要がある。その為には、解釈と制約から例外データを検出するための基本的手段をシステムが持つ必要があり、ここでは、入力される主張間の論理的整合性を計算するものであり、記述力に優れ、主張の追加／削除が容易である McAllester のＴＭＳ[2][3]の利用[4]を検討した。

ＤＢからの知識抽出方法をシステムとして構成した場合の概念的な機能フローを図２に示す。



（図２．ＤＢからの知識抽出システムの機能フロー）

Automatic Knowledge Acquisition from A Database that Includes Exceptional Cases
Takefumi YAMAZAKI,Satoshi KUWAHARA,Fumio HATTORI
NTT Communications and Information Processing Laboratories

処理の流れは、以下の通りである。

まず、データから生成された解釈／制約条件がＴＭＳに入力されると、ＴＭＳはその論理的整合性を計算し、例外データ候補を提示する。次に、それら例外データ候補の中から、出現頻度に基づいて、例外データを選択して、知識抽出の対象から除去し、矛盾を解消する。最後に、組合せの整合性をＴＭＳでCHECKしながら、整合性のある解釈の組合せを知識として、蓄積する。

4．実験によるシステムの評価

前章までの知識抽出モデルを評価するため、化学反応ＤＢからの化学知識抽出を評価の題材に設定して、簡単なプロトタイプを作成した。　具体的には、既知の知識ではあるが、ベンゼン環における置換基の位置選択性及びそれらの相対的強度に関する知識を既存の化学反応ＤＢから自動的に抽出することを試みた。実験に使用した反応ＤＢ〔5〕〔6〕は、主に文献〔7〕より収集されたもので,約７０００件の反応事例が格納されていた。それから今回の知識抽出に必要なデータ（ベンゼン環の置換反応）を約１００個検索し、それらのデータを用いて、置換基の位置選択性、及び相対的強度に関する知識の抽出実験を行なった。

用いたデータ・そのデータから生成される解釈・制約条件の例を図に示す。これらの解釈／制約条件を図３に従って、本抽出システムに投入した結果の出力を、図４に示す。

```
データ：(AR-OH O AR-OCH3 M)
解釈１：(AR-OH O)∧(AR-OCH3 M)
解釈２：(AR-OH O)∧(AR-OCH3 O)∧(AR-OH〉AR-OCH3)
解釈３：(AR-OH M)∧(AR-OCH3 M)∧(AR-OH〈AR-OCH3)
制約条件１：(AR-OH O)∧(AR-OH M) --〉 ⊥
制約条件２：(AR-OCH3 O)∧(AR-OCH3 M) --〉 ⊥
制約条件３：(AR-OH〉AR-OCH3)∧(AR-OH〈AR-OCH3) --〉⊥
```

（図３．データ・解釈・制約条件の例）

```
******* RESULT OUTPUT*****
*** O ***
 AR-OCOCH3-is-O = TRUE
 AR-OCH3-is-O = TRUE
 AR-N(CH3)2-is-O = TRUE
 AR-CL-is-O = TRUE
 AR-OH-is-O = TRUE
 AR-BR-is-O = TRUE
 AR-NHCOCH3-is-O = TRUE
 AR-CH3-is-O = TRUE
 AR-NH2-is-O = TRUE
 AR-NO2-is-O = TRUE

*** M ***
 AR-COCH3-is-M = TRUE
 AR-CHO-is-M = TRUE
 AR-COOCH3-is-M = TRUE
 AR-COOH-is-M = TRUE

*** relative strength ***
 AR-OH〉AR-OCH3 = TRUE
 AR-OH〉AR-CL = TRUE
 AR-OH〉AR-NO2 = TRUE
 AR-OH〉AR-BR = TRUE
 AR-NH2〉AR-NO2 = TRUE
 AR-NH2〉AR-COOH = TRUE
 AR-OH〉AR-CH3 = TRUE
 AR-NHCOCH3〉AR-CH3 = TRUE
 AR-NH2〉AR-CH3 = TRUE
 AR-CH3〉AR-NO2 = TRUE
 AR-CH3〈AR-NH2 = TRUE
```

（図４．知識抽出の結果）

置換基の配向性の知識は化学上の既知の知識であることから、図４の結果の内、配向性に関しては、参考文献〔8〕により、AR-NO2 以外は、その一致が確認できた。AR-NO2 の場合は、たまたま特殊な反応データのみであったため、理論と一致しなかったと考えられる。

また、図４の結果の相対的強度に関しては、化学専門家に検証依頼した結果、正しいことが確認できた。

5．まとめ

本研究では、ＤＢ内の事例データから得られる解釈を組み合わせて、整合性のある解釈の組合せを知識として生成することを基本とするＤＢからの知識自動抽出方法を提案した。その際の例外データの検出にはＴＭＳを用い、かつ例外事例の除去には事例頻度に基づく判定法を採用することにより、整合性のある解釈の組合せとして、知識を抽出するシステムが構成できることを示した。

本システムの有効性については、化学反応ＤＢからベンゼン環での置換基の位置選択性知識の抽出実験を行なうことにより、確認を行なった。

今後の課題として、解釈生成のための知識の獲得や出現頻度以外の例外データ同定ための評価方法の検討が考えられる。

6．謝辞

本研究に用いた化学反応ＤＢ、ＤＢ利用のソフト（反応検索システム、反応箇所抽出システム）を試用させて頂いた中外製薬株式会社富士御殿場研究所の松浦育敏氏、及び獲得した化学知識の検証に御協力頂いた名古屋大学化学測定機器センターの早川芳宏助教授に感謝致します。

また、本研究の機会を与えて頂いた村上知識処理研究部部長、吉田主幹研究員に感謝すると共に、熱心に御討論頂いた、知識処理研究部の方々に感謝致します。

7．参考文献

〔1〕山崎　毅文、桑原　敏：”KBMSによる化学反応知識構成法”　　情処第３７回全大，1988

〔2〕McAllester,D.：”An outlook on truth maintenance”,Artificial Intelligence Laboratory,AIM-551,MIT,Cambridge,MA,1980

〔3〕McAllester,D.：”A three-valued truth maintenance system,S.B. Thesis "Department of Computer Science,Tech.Rept.No.203,State University of New York, Buffalo,NY,1983

〔4〕山崎　毅文、桑原　敏、服部　文夫：”例外事例を含むＤＢからの化学知識の自動抽出”，情処第６５回研究会，1989

〔5〕松浦　育敏：”データベース型反応設計支援システム”，知識システムによる分子設計研究会，1986-9

〔6〕松浦　育敏：”化学物質等設計知識ベース研究の方針と現状”知識システムによる分子設計研究会，1988-9

〔7〕Dauben G.William , ed. : "ORGANIC SYNTHESES" ,Wiley ,1963

〔8〕Hugh J. Williams：”入門・有機化学”化学同人，1987