

TM-0397

類推タスク構造に基づく知識獲得
—分類型問題向け知識獲得支援システム CTAS—

山崎毅文 (NTT)
瀧 寛和, 椿 和弘

October, 1987

©1987, ICOT

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03) 456-3191~5
Telex ICOT J32964

Institute for New Generation Computer Technology

類型タスク構造に基づく知識獲得

Knowledge Acquisition based on Generic Task Structure

— 分類型問題向け 知識獲得支援システムCTAS —
(Classification Task Acquisition System)

山崎 毅文 (N T T情報通信処理研究所) 滝 寛和、椿 和弘 (I C O T)

エキスパートシステムにおける重要なタスクの一つである分類タスクは、その初期段階においては階層分類タスク、その末期段階においては順位付けタスクからなっていると考えられる。本研究において、この分類型タスク構造を考慮した分類型問題向け知識獲得支援システムCTASをP S Iマシン上に構築した。

1. はじめに

知識獲得は、知識システム構築における最も重要な作業であり、それは、専門家の知識を効率よく収集し、整理・体系化を行い、エキスパートシステム構築ツールの知識表現に変換する作業である。しかしながら、知識獲得は、専らKE (ナレッジエンジニア) の手作業にたよっているのが現状である。というのは、現在エキスパートシステム構築に対する組織的方法論はなく、構築に際して用意されているのは、エキスパートシステム構築ツールのみである。その構築ツールも、記述レベルが低すぎるため、KEの負担が、非常に大きく、またその構築支援能力にも限界があると考えられる。

知識ベースの早期構築や信頼度の高い知識ベースの構築のためには、エキスパート自らが使用して簡単に知識を投入し、その実行結果や推論過程を簡単に確認できるような、知識獲得支援システムが必要である。

知識獲得支援システムを構築する上で、エキスパートシステムが行なう知識情報処理タスクのタスク構造に注目して考えるのは有効であろう。そこで、そのタスク構造の一つである分類タスクについての知識獲得支援を考えた。分類タスクは、階層分類タスクと順位付けタスクの2つの類型タスク (GENERIC TASKS) からなり、分類の初期段階では階層分類を、末期段階では特徴による順位付けを用いて、分類を行なっている。

今回試作した分類型問題の知識獲得支援システムCTAS (Classification Task Acquisition System) は、この分類型タスク構造を考慮して構築したシステムである。

2 分類型問題タスクとは?

2-1 分類型問題

エキスパートシステムは、その質的構造により、診断問題等の分類型問題と設計問題等の合成型問題とに大別される。このうち、分類型問題の類型タスク構造は、ある程度明確になりつつある。そのタスク構造は、階層分類 (Hierarchical Classification) と特徴による順位付け (Ordering by Specification) とから成ると考えられる。例えば、MYCINのようなシステムは、病名の分類をCF値による順位付けタスクによって行っていると考えられるし、また機械の故障診断のような症状と故障原因とが一意に結びつくようなシステムは、診断木の探索によって解を探すタスクと考えられよう。また、一般的に、分類の初期段階では階層分類を、末期段階では順位付けタスクが多い。分類活動の初期段階では、分類すべき項目数が多いために、項目間には項目をグループ分割できるような特徴 (特徴1と呼ぶ) が多く、その特徴で分類対象をグループ分割す

ることができる。一方、分類が進んだ末期段階では、性質の似たようなものを分類しようとしているので、もはやそれらをグループ分割できるような特徴は存在しない。これらを分類するのに必要な特徴は、特徴1とは別の特徴でなければならない。この特徴を特徴2としよう。特徴2は、分類対象各項目と何らかの関係を持っているが、項目と特徴2との関係度の強さが各項目によって異なる。このことを利用して分類を行うのである。これは、各項目を特徴2によって順位付けすることによる分類タスクといえる。

階層分類タスクと順位付けタスクについて、次の章で述べる。

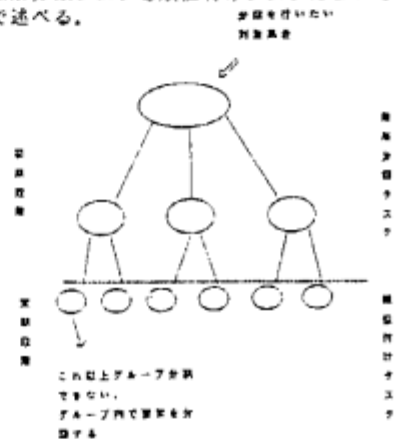


図1. 分類の初期段階と末期段階

2-2 分類の様相

(1) 階層分類型タスク向き

分類の初期段階に見られる傾向で、医療診断の場合、診断の初期に行なう主訴別分類タスクにあたる。ある特徴で、集合要素をほぼ分けることができる。



図2. 階層分類タスクの例

(2) 順位付タスク向き

分類の末期段階に見られる傾向で、ある特徴で、集合要素を分けることができないタスク

要素AもBも特徴a, b, cを持っているときに、どういうやり方で、AとBを分類すればよいだろうか。「Aとa, b, cの関係の強さ」と「Bとa, b, cの関係の強さ」からどちらがよりa, b, cの関係を持つかによって、AとBを分類するしかない。

例えば、以下のような知識ベースが与えられているとする。

A	B	C	要素/特徴
5	2	1	a (5...1)
2	5	4	b (5...1)
3	4	1	c (5...1)

図3. 簡単な知識ベースの例

今、特徴a, b, cについて、観測された値が、4, 4, 4とすると、

Aに対するベクタ (5 2 3)

Bに対するベクタ (2 5 4)

Cに対するベクタ (1 4 1) 以上の各ベクタと、今観測された値からなるベクタ (4 4 4) とのノルムによって、A, B, Cどれが、今観測された状態に近いかを判断する。

医療診断の場合だと、要素(A, B, C)にくるのが病名で、特徴(a, b, c)にくるのが、症状と考えればよい。

3 分類型知識獲得支援システムCTAS

本システムは、ICOTで開発されたPSIマシン上に構築されている。PSIのOSであるSIMPOSは、ウィンドウやオブジェクト指向ロジックプログラミング言語E.S.P.開発環境等の豊富な機能が利用でき、知識獲得を効率良く行うために必要なユーザインターフェースを構築することができた。使用言語は、E.S.P.である。

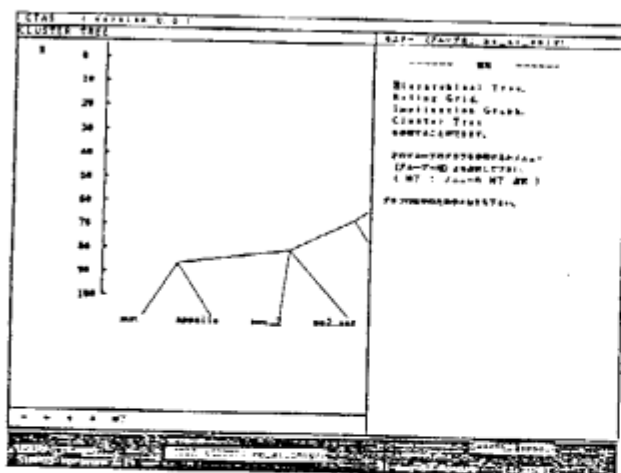


図4. CTASの全体画面

知識獲得フローは、次の4段階で行われる。

- (1) 抽出
- (2) 整理
- (3) リファイン
- (4) テスト

3-1 抽出プロセス

先の分類の様相を考えて、2通りの知識抽出法を、CTASでは提供している。

(A) ボトムアップ式抽出法

(1)まず、分類したい対象集合の要素項目(ELEMENT)を抽出し、

(2)その後、グループがBINARY TREEになるように、ELEMENTのグループ分割(階層分類)を行なう。

以上で、階層分類は終了し、次に順位付タスクを行なう。

(3)各のグループ内のELEMENTの分類を行う。その分類に必要な特徴(TRAIT)と関係度の抽出を行なう。まず、特徴抽出を行なうが、その方法は、ETSで行われているPersonal Construct Theoryを用いた特徴抽出法を適用する。

(4)次に、ELEMENTとTRAIT間の関係度の抽出を行う。すると、グループ毎に、ELEMENTとTRAITの関係表(RATING GRID)ができる。

以上で、ELEMENTの階層分類とグループ毎のRATING GRIDが作成されたことになる。

E1	E2	E3	...	element / trait
5	4	1		c1/c1'
1	5	5		c2/c2'

図5. RATING GRIDの例

(B) トップダウン式抽出法

ELEMENTを階層分類した時の階層構造が、予め解っている時にこの方法が有効である。

(1)まず、最初に、階層構造の枠組みをBINARY TREEで抽出する。(A)でいうグループを先に抽出することに当たる。

(2)(1)によって、できた各グループに属するELEMENTを抽出する。

以下の操作(3),(4)については、先の(A)と同じである。

セミナー

----- 質問 -----

月についての分類をするのですか。

分類を >> machine

要素抽出から先にしますか? (y/n) y

*** 要素抽出 ***

要素を1つずつ入力して下さい。

(要素入力終了 : and)

要素抽出 >> 3700000000

要素抽出 >> 4000000000

要素抽出 >> 4000000000

要素抽出 >>

図6. ボトムアップ式抽出法 (次頁に続く)

```

*** 属性分類 セード ***

属性を分類する対象を入力して下さい。

グループ分け >> personal_unix

その属性の情報を入力して下さい。

グループ分け >> multi_unix

そのノニュームグループ分けを
知ってください。

*** 加工処理 セード ***

グループの行名の結合を行います。
personal_unixとmulti_unixの行名結合を
行います。

multi_unix, u_station, pc2_unix
のうちで2つに共通する行名は ??

印刷処理 >> business_exit

行名の情報を入力して下さい。

印刷処理 >> no_business_exit

u_station, pc2_unix, sun のうちで2
つに共通する行名は ??

印刷処理 >>

*** rating 入力 ***

グループ毎にrating値の入力
をお願いします。
グループ名がpersonal_unixとmulti_unixである行
名について
rating値のを入力してください。
business_exit_unix(5)... no_supe
r_unix(1)
(symbols) ** 5

(c1-explorer) **

```

図 6. ボトムアップ式抽出法

3-2 整理プロセス

グループ間の関係を表す階層木(HIERARCHICAL TREE)やグループ毎のRATING GRIDを作成する。また、rating gridを基に、implication graph(trait間の関係を示すグラフ)、similarity表示(trait間、及びelement間の類似度を示す)、cluster tree(trait間・element間の類似度を示すグラフ)を作成する。これらの表・グラフによって、ユーザは、知識の内容を視覚的に捉えることができ、知識ベースのリファインを行いやすくしている。

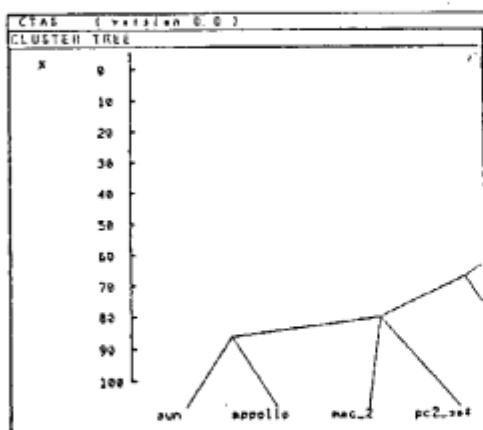


図 7. cluster tree (element間)

3-3 リファインプロセス

先の整理プロセスで表示した表・グラフを基に、階層木の修正、rating gridの値の修正、新たなtraitやelementの追加等を行って、知識をリファインする。

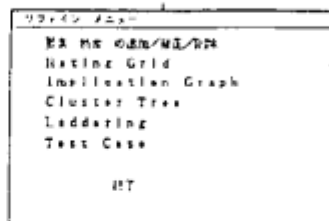


図 8. リファインメニュー



図 9. cluster treeによるリファイン

隔層水・rating gridを基にして、次の3種類のルールを作成する。

(8) CONCLUSION RULE (各特徴と結論項目間の関係を示すルールで、CF値付きルールとなる)

(C) INTERMEDIATE RULE (各特徴間の関係を示すルール。IMPLICATION GRAPHに表れる特徴間についてのみ、作成される。)

作成されたルールは、PSI-PROLOG上のCTAS用推論エンジンで実行される。まず、Hierarchical ruleの実行が行われ、テストの対象になっている結論項目の実行するグループが決定される。次にそのグループに属する結論項目と特徴に關するConclusion RuleとIntermediate Ruleが実行される。これは、該グループに属する特徴の確信度に対する閾値に対して、-1.0~1.0の値を入力することによって行われる。これらのルールの実行後、結論項目が、確信度の高い順にソートされて表示され、満足のおく結果が得られない時は、リファインモードに戻ってCTASの提供する種々の手法でリファインを行うことになる。

```

CTAS >> test,

Element group is set

Please input pos or no

var_environment is 'var_env' or multi_var (no) ?
var_env is 'elem_only (pos) or no_elem_only (no) ?
>> no,

This state name is var_env+no_elem_only

Group definition is over

Next, elements are set in order of CF value

Please input CF value (-1, 0-1, 0)

test is 'var_env (-1, 0) / no_elem_only (-1, 0) / unknown (0, 0) ?
>> 1, 0,
test is 'standard_elem (-1, 0) / no_standard_elem (-1, 0) / unknown (0, 0) ?
>> 0, 0,
mode_elem is 'mode_in_elem (-1, 0) / no_mode_in_elem (-1, 0) / unknown (0, 0) ?
>> 0, 0,
test is 'no_heterogeneous (-1, 0) / heterogeneous (-1, 0) / unknown (0, 0) ?
>> 1, 0,
mode is 'multi_window (-1, 0) / no_multi_window (-1, 0) / unknown (0, 0) ?
>> 1, 0,

The result of test is below

machine is 'max_2 (0. 99922451)
machine is 'x2_value (0. 875)
machine is 'xvalue (0. 5)
machine is 'var (0. 1230405)
machine is 'var_env (-0. 0375) -bit
machine is 'where_var (0. 4290875)
machine is 'w_value (0. 546875)

```

図 10 テストの例

本システムでは、以上の抽出・整理・リファイン・テストを繰り返すことにより、知識ベースを構築していく。知識獲得は、システムとユーザとがインタラクティブに処理を進めて行くことにより行なわれる。

4 主とめ

分類タスクのタスク構造を考慮して構築した知識獲得支援システムCTASについて報告した。

CTASは、分類型問題に関する知識獲得について強力なツールであるが、次の2点について未解決な問題を残している。

1つは、ボトムアップ、トップダウンの階層分類は直感的には有効であると思われるが、理論的根拠に乏しいことである。Personal Construct Theoryがあったが、最初の階層分類における特徴抽出では

決定的方法がない。もう一点は、階層分類と特徴順位付けとの境目をどこでつけるかという問題である。分類活動は、この2種類のタスクから成り立っているのは、明かであるが、その境界ははっきりしていないように思われる。これら2点の問題解決が、今後の検討課題である。

今後の研究、展開として、次の4点を挙げるこ
とができる。

(1) CTAS の評価

CTASを使って知識ベースを作成していく過程で、どの技術が、知識獲得プロセスで重要なのかを説明する。そして、知識獲得のための要索技術として切り出せる部分があるかどうかを検討する。

(2) 他の知識獲得モデルとの結合

CITASは、エキスパートシステムの対象問題の類型タスク構造を基に構築した知識獲得支援システムであるが、他の知識獲得モデルと結合することにより、より強力な知識獲得ツールができるであろう。例えば、MOREシステムにおける知識リファイン方法、KJ法による知識抽出法、ICOTで開発中のEPSILON/EMシステムのトップダウン式知識獲得法等を参考に、利用できる方法をCITASに取り入れることを検討中である。

(3) 心理学的アプローチの研究

CTASは、特徴による順位付けにETSシステムのやり方を一部採用しているが、ETSにおいては、様々な心理学的手法が用いられている。その個々の手法についての有効性を調べたり、また計量心理学や専門家に対するインタビュートクニク等の他の心理学的手法を調べることは、有効であろう。

(4) ユーザインタフェースの研究

ユーザが使いやすいインターフェースは、どんなものか。ユーザにとって簡単に入力でき、また知識の内容も見やすく、リファインもしやすい知識表現形式は、何かを検討する。

[附 录]

本研究の機会を与えて下さったICOT研究所蔵一博所長並びにNTT情報通信処理研究所堀内主幹研究員に感謝致します。また、本システム構築に御尽力頂いた大崎氏(JIPDEC)、並びに熱心に討論して頂いた藤井室長をはじめとする第五研究室の諸氏に感謝致します。

【参考文献】

- [Boose 85] Boose, J. : Personal construct theory and the transfer of human expertise, Advances in Artificial Intelligence., North-Holland
[Chandrasekaran 85] Chandrasekaran, R. : Generic Tasks in Knowledge-Based Reasoning: High-Level Building Blocks for Expert System Design., IEEE Expert
[Chikayama 84] Chikayama, T. : Unique Features of ESP, International Conference on Fifth generation Computer Systems
[ICOT 87] ICOT Technical Report, "Knowledge Base for Expert shell Evaluation
[Kahn 85] Kahn, G., Nowlap, S., McDermott, J. : MOR E: An Intelligent Knowledge Acquisition Tool, Proc. of IJCAI'85
[篠原 87] 篠原 健志 : 未整理な情報からの知識ベース構築, 情処学会第34回全国大会
[滝 87] 滝 寛和, 他 : 知識獲得支援システム (E P S I L O N) における専門家モデル, 情処学会知識工學と人工知能研究会報告NO.52