

教示すべき内容を指示する機能を持つ
学習システム構築の試み

3E-9

田代 勤、 明石 吉三

(株)日立製作所 システム開発研究所

1. はじめに

ルール型制御、故障・医療診断等のルール型の知識処理システムが多数開発され効果を上げている。しかし、大規模な対象では、矛盾がなく質の良いルールを作成するのに多くの工数を要する。この工数を削減するための方法として、機械学習の手法を用い、対象状態と結論の対からなる具体例のみを与え多くの対象状態に適用できるルールを自動的に導くルール獲得方式が注目されている。

このルール獲得方式では、対象状態を事例記述に、結論を概念名に、学習終了時の概念記述をルールの条件に対応させ、ルール獲得の問題を概念学習の問題に置き換えて扱う。ここで、従来の概念学習のシステムは、事例の与え方を教示者にまかせている。このため、学習システムから見て偏っていたり、無駄な事例が与えられる場合があり、なかなか学習が終了せず多くの事例を必要とするという問題があった。本稿では、学習システム側から教示者に必要十分な事例を入力できるよう指示する一方法を提案する。

2. 概念学習問題の例

初めに概念学習問題の例を説明しておく(図1)。学習すべき概念の名前はA~Dである。各概念の具体的表現(事例記述)及び一般的表現(概念記述)は2つの属性の組からなる事例、概念記述言語により記述される。それぞれの属性値は階層関係を持ち、上位ほど下位の値を含む一般値となっている。[[x3, y1], C]のような事例記述とその概念名の対からなる事例を得て、学習目標に示すような複数の事例記述を含む概念記述(例えば、Dの概念記述 [X1, Y2])を各概念に対して見い出すことが問題となる。

3. 入力事例指示方式

3.1 事例指示型学習システムの枠組

上記学習問題に対し新アルゴリズム(複合多重集約アルゴリズム¹⁾)を提案している。本稿ではこのアルゴリズムを前提に、次に入力すべき事例を指示する方法を提案する。事例指示型学習システムの枠組を図2に示す。複合多重集約アルゴリズムでは、概念毎に、与えられた事例を全て包含する最も一般的な概念記述の集合を上限、最も特殊な概念記述の集合を下限として保持(上限と下限の間をバージョン空間と呼ぶ)している。事例記述選択アルゴリズムは、それぞれのバージョン空間の上、下限の要素間の関係を解析し次に入力すべき事例の実例記述を提示する。教示者は、その事例記述に対応する概念

学習概念: {A, B, C, D}
 事例、概念記述言語: [属性x, 属性y]
 $x \in \{x1, x2, x3, x4, X1, X2, *\}$
 $y \in \{y1, y2, y3, y4, Y1, Y2, *\}$
 事例記述は__のみ使用

属性値の一般化関係:

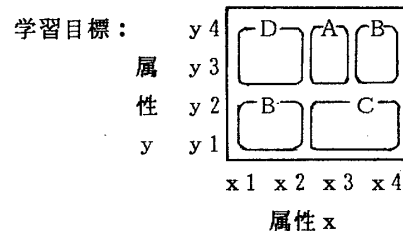
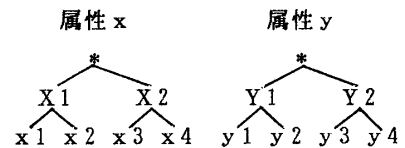


図1 概念学習問題

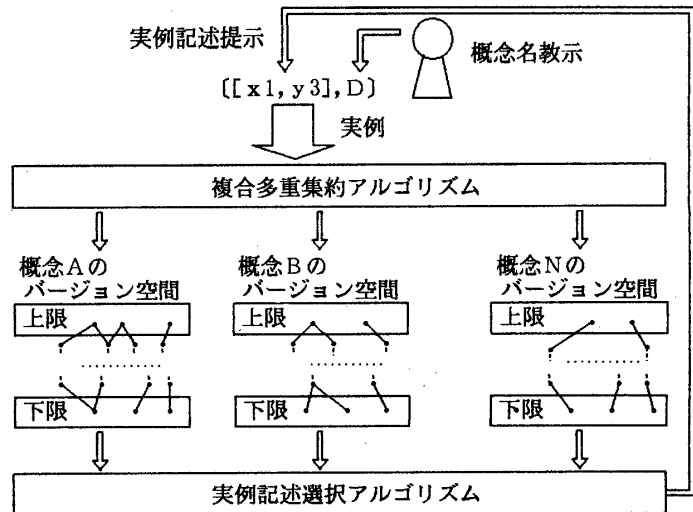


図2 事例指示型学習システム

の名前を与え実例として入力する。以上のサイクルを繰り返し全バージョン空間の上、下限が一致すると学習を終了する。

3.2 実例記述選択方法

実例指示の目的は、バージョン空間を速く収束できる実例を得て、少ない実例で学習を終了可能とする点にある。提案する実例記述選択法は、実例としてある実例記述を得た場合、どの程度バージョン空間が収束するかかの尺度に基づく方法である。まだ実例として入力されていない実例記述毎にこの尺度を計算し、最も大きい値を持つ実例記述を、次に入力すべき実例の実例記述として提示する。

複合多重集約アルゴリズムにおいて、各概念のバージョン空間の上、下限及び与える実例間の関係を、学習の前と後のバージョン空間の縮退度の大きい順に整理して表1に示す。複合多重集約アルゴリズムは、ある概念の実例が与えられると、その概念の下限要素を与えられた実例の実例記述を包含するように拡大する。更に、拡大された下限要素とオーバーラップしないように他の概念の上限要素を縮小する。なお、与えられた実例の下限要素がまだ存在しなければ、与えられた実例の実例記述を下限要素としてその実例の概念の下限に登録する。これらのことから、表1に示したように、多くの概念の上限要素に共通に含まれかつ含まれる上限要素に更に下限要素が存在する実例記述ほど、バージョン空間の縮退度(上限と下限の差の減少度)が大きい。よって、以下を実例記述選択の尺度とする。

上限分離度：実例記述が含まれる上限要素の数。これが大きいほど、実例として得た場合縮小される他の概念の上限要素が多い。

下限拡大度：上記上限要素のうち下限要素を持つもの数。これが大きいほど、実例として得た場合下限要素が拡大される確率が高く、他の概念の上限要素が更に縮小される可能性が大きい。

上限分離度が最も大きく、これが同じならば、更に、下限拡大度の最も大きい実例記述を選択する。これらの尺度にて一意に定まらなければランダムに選択する。

4. 適用事例

図1の問題に対し実例をランダムに入力した場合と提案方式した実例指示方式による場合の一学習例を図3に示しておく。学習過程の詳細は省略するが、提案方式では、ある概念の実例として既に推定されている実例を更に入力するといった無駄や下限要素の拡張度の低い実例を入力するといった事態が避けられている。これらから、ランダム入力に比べ2/3の実例入力力で学習を終了しており提案方式の効果が期待できる。

5. おわりに

機械学習手法の実用性を高めるべく少ない実例で学習を完了できる入力実例指示方式を提案した。更に、多くの例を用いて評価、改良を行なっていく。

参考文献：1) 田代他：複数概念の選言表現の逐次的学習のための複合多重集約アルゴリズム，情報処理学会論文誌，30-9，pp.1073-1082(1989)

表1 与えた実例とバージョン空間の縮退度の関係

学習前の上下限要素の関係	学習後の上下限要素の関係	空間縮退度
		大 ↑ ↓ 小

○, ○, ○ : 概念A, B, Cの上限要素
 ◐ : 概念Aの下限要素 ○ (dashed) : 与えた概念Aの実例

ランダム入力：

- [[x4, y2], C], [[x3, y3], A], [[x4, y3], B],
- [[x1, y1], B], [[x3, y2], C], [[x1, y4], D],
- [[x1, y2], B], [[x3, y4], A], [[x3, y1], C],
- [[x2, y2], B], [[x4, y1], C], [[x1, y3], D],
- [[x2, y1], B], [[x4, y4], B], [[x2, y4], D]

指示入力：

- [[x4, y2], C], [[x1, y1], B], [[x3, y3], A],
- [[x2, y4], D], [[x3, y1], C], [[x2, y2], B],
- [[x4, y4], B], [[x1, y3], D], [[x4, y3], B],
- [[x3, y4], A]

図3 学習例の比較