

4 L - 9

# 仮説探索の三次元視覚化機能を備えた ILP システムの設計\*

吉井 崇行<sup>†</sup> 大和田 勇人<sup>†</sup> 溝口文雄<sup>†</sup>東京理科大学 理工学部 経営工学科<sup>†</sup>

## 1 はじめに

既存のILP(帰納論理プログラミング, Inductive Logic Programming)による学習システムは, 仮説探索において探索空間が視覚化されていないため, ユーザは得られたルールがどのように導かれたのかが理解しづらい。

そこで, 本稿では, ルール導出の過程を視覚化できるツールを設計する。視覚化は, 三次元で行なう。なぜなら, 三次元は, 1) 二次元よりも扱える情報が多く, 2) 奥行きを利用した対象物の表示, ユーザが視点切替えによる任意の場所から眺望できる, といった特徴があるからである。そのため, 述語間の関係, 仮説間の関係を効果的に見せることが可能であると思われる。

## 2 設計方針

本ILPシステムは, 大きく分けると, 学習部と視覚化部から成る。

学習部は, Muggleton の逆伴意 (Inverse Entailment)[1]と溝口らによって開発されたGKS[2]を参考にして作成した。本学習システムの特徴は, 正事例数とルールの長さを目的関数とし, 負事例を含んでもよい割合を制約条件として与え, 生成されたボトム節の中から最も良い仮説を見つけることである。

視覚化部に関しては, 第一階述語論理で記述でき, 関係学習が可能であるILPシステムの特徴を活かせるように, 述語論理で記述された探索された仮説の視覚化, 探索された仮説が持つ数値情報の視覚化を考える。そこで, 学習システムが出力した探索履歴ファイルをもとに, 3種類の三次元視覚化ツールとして, コーンツリー, ランドスケープ, そして, キューブを作成する。

コーンツリーは, 探索された仮説の視覚化を行ない, ランドスケープは仮説の重要度を表現する。そして, キューブは, 探索された仮説が持つ数値情報の分布情報を視覚化する。

また, 探索履歴ファイルを取得するのは, 学習部と視覚化部をリアルタイムに接続することを避けるためである。一つの理由として, 短時間で学習が終るのであればよいが, ルールが求まるまで非常に時間がかかる場合, ユーザはその間ずっと表示対象を見ておかなければならず, 現実的ではないことが挙げられる。そこで, 本稿では, 学習部が出力した探索履歴ファイルを視覚化機能が読み込み, 再生する方法を採用する。

実装は, Javaで行なう。また, 三次元視覚化は, Java3Dで実装する。

## 3 探索履歴ファイルの作成方法

本システムの全体図を図1に示す。まず, データファイルを読み込み, 学習し, ルールが得られる。その際, 仮説の探索情報は, 探索履歴ファイルとして出力される。そして, その履歴ファイルを履歴オブジェクトとして扱い, コーンツリークラス, ランドスケープクラス, キューブクラスに伝達し, 表示を行なう。

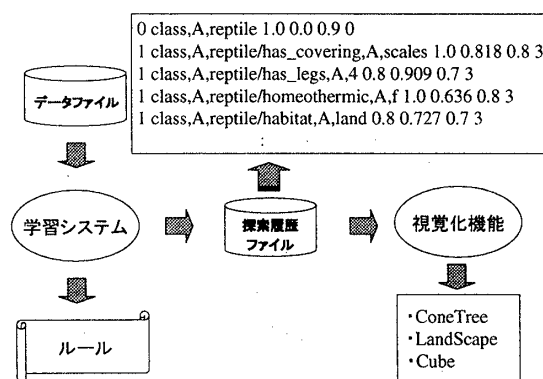


図1: システムの全体図

探索履歴ファイルの各行は, それぞれの探索された仮説に対して, 1) 節の長さ, 2) 仮説, 3) 全正事例に対して仮説が包含可能な正事例の割合 (Sensitivity), 4) 全負事例に対して, 仮説に包含されなかった負事例の割合 (Specificity), 5) 節自身が持つ評価値, 6) 探索情報, のような履歴情報が書かれている。

\*Design of an ILP system with 3-dimensional visualization of hypothesis searching

<sup>†</sup>Takayuki Yoshii, Hyato Ohwada, Fumio MIZOGUCHI

<sup>†</sup>Department of Industrial Administration, Faculty of Sci. and Tech., Science University of Tokyo

## 4 視覚化機能

三次元視覚化ツールは、3章に書かれた方法で作成された探索履歴ファイルを読み込み、コーンツリー、ランドスケープ、そして、キューブの表示を行なう。以下には、各々の視覚化ツールの目的と実行例を示す。

### ● コーンツリー (図2)

コーンツリー [3] は大規模データを効率的に収納でき、同一階層と階層間の関係を表現できる。それゆえ、長い節、そして、大規模な述語の組み合わせも考慮に入れた視覚化が可能である。各階層には節を構成する述語があり、各述語名には色が付いていて、探索情報を表現している。述語名の色は、探索履歴ファイルの探索情報によって決まる。

利点： 仮説間の関係を把握することができる。また、どのような位置にルールかがあるかわかる。

### ● ランドスケープ (図3)

ルールとは、正事例の包含数が多ければ多いほどよく、逆に、負事例の包含数が少なければ少ないほどよい。そこで、それが反映されている数値情報である Sensitivity と Specificity の値に注目する。各階層には節を構成する述語があり、Sensitivity と Specificity の2種類の数値情報を高さで表現した。

利点： 各仮説の重要度を視覚的に捉えることができ、ルールがどのような場所にあるかが理解できる。

### ● キューブ (図4)

探索された仮説は、Sensitivity, Specificity, そして、探索時の評価尺度といった3種類の値を持つ。そこで、Sensitivity, Specificity, 評価値を表現する3つの軸を作成し、履歴情報ファイルから探索された仮説のそれら3つの値と探索情報を読み取ってキューブ中に点(球)をプロットした。プロットした点の色は、履歴情報ファイルの探索情報によって決まる。また、本学習システムは、指定された特殊化率(Specificity)を制約として探索を行なう。その特殊化率をキューブ中に透明な壁として表現した。

利点： 仮説全体の傾向や、制約条件と仮説の関係を把握できる。また、ルールが全体のどの辺りにあるかも視覚的に理解できる。

## 5 まとめと今後の課題

ILPシステムのルール導出過程を明らかにするため、3種類の三次元視覚化ツールを作成した。しかし、大規模な仮説探索を行なった場合、履歴情報もそれにつれて増える。表示に必要な情報を絞り込んだとしても、完全に対処しきれないと考えられる。そこで、インターフェースを作成し、ユーザのニーズに合わせたカスタマ

イズを行なう必要がある。また、本システムをユーザに利用してもらい、評価を行なう予定である。

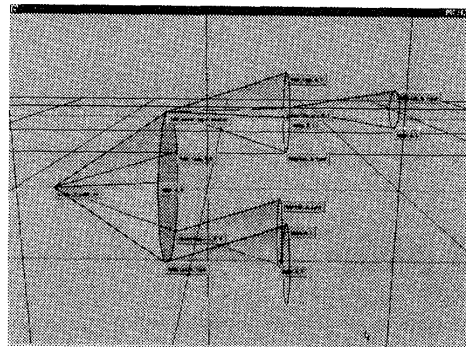


図2: コーンツリー

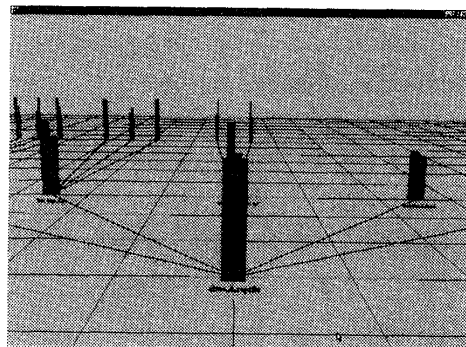


図3: ランドスケープ

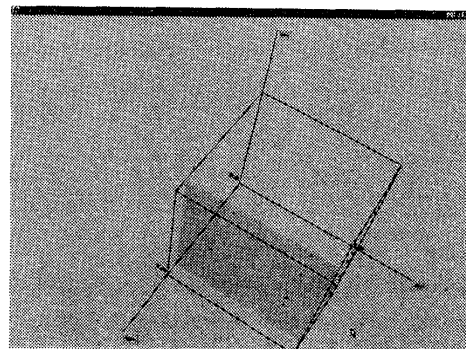


図4: キューブ

## 参考文献

- [1] S. Muggleton. 'Inverse Entailment and Progol', *New Generation Computing*,13:245-286, 1995.
- [2] F.Mizoguchi and H.Ohwada. Using Inductive Logic Programming for Constraint Acquisition in Constraint-based Problem Solving. Proc. of the 5 th International Workshop on ILP, 297-322,1995.
- [3] Hideki Koike, Hirotaaka Yoshihara, Fractal Approaches for Visualizing Huge Hierarchies, Proceedings of 1993 IEEE Symposium on Visual Languages (VL'93), pp.55-60,1993.