関係型データマイニングにおけるパターンの視覚化表現

堀 弘樹 （指導教員：犬塚 信博 教授）
名古屋工業大学 工学部

１．はじめに

データマイニングとは，大量のデータから隠された知識や新しい規則を発見するプロセスである．その中でも**関係型データマイニング**(以下 MRDM)は**述語論理**で表現されており，比較的ユーザにとっても理解し易い形式となっているが，大規模データベースを扱う場合には理解し難くなってしまうため，理解し易い出力形式が必要となる．既存の形式としては，当研究室吉野らによって提案された視覚化表現[1]があるが，生成されたルールにはそれでも表現できない構造や，ルールごとに違った意味を持つ構造が存在する．例えば，人物同士のつながりを表す**接続関係**や，全体がどの部分を含むかを表す**包含関係**である．本研究ではこれらの意味を考慮した視覚化表現を提案する．

２．既存手法の問題点

例えば，以下のようなパターン p1，p2 があるとする．

p1:grandfather(koji) ← parent(koji, yozo) ∧
parent(yozo, kyoichi) ∧ male(kyoichi)
p2:train(t) ← has_car(t, c) ∧ has_load(c, l) ∧
triangle(l) ∧ load_num(l, 2)

p1，p2 を視覚化表現に変換した際，理想的な表現は図１，図２となるが，パターンの種類の膨大さや機械的に処理することを考えると現実的ではない．そこで吉野らによって図３のような有向グラフ表現が提案された．しかし吉野らによる手法は頂点と辺が元々の述語の項と述語名から成る単純なものであり，視覚化表現を用いても構造が理解し難い．また，包含関係を表す視覚化表現(p2)が表現できない．

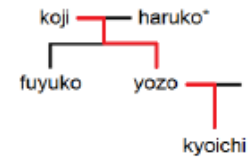

図１:p1 の視覚化

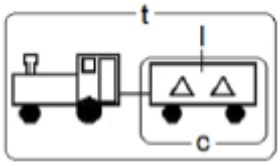

図２:p2 の視覚化

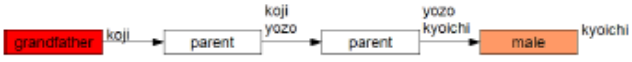

図３:吉野らによる手法

３．提案手法

パターンは以下のような 3 つの対象から成ると想定し，それらを分類することによりパターンを視覚化表現に変換することを考える．

個体・・・具体的な対象．
属性・・・個体を修飾する抽象的な対象(個体の性質)．
関係・・・個体同士を関連付ける対象．

ここで，一つの項からもう一つの項を関数的に呼び出す働きをする**モード**を利用して，述語の述語名と項から上記 3 つの構成条件を次に与える(＋は入力引数，－は出力引数，＃は定数引数を表す)．

個体・・・入力引数，出力引数．
属性・・・定数引数，出力引数のみを持つ述語の述語名，定数引数を持つ述語の述語名．
ただし，前者 2 つを特に属性頂点，後者を特に属性辺とする．
関係・・・出力引数を持つ述語の述語名．

上記の条件のもと p2 について調べてみると，表１のようになる．グラフの考えをもとに個体と属性頂点を**頂点**，属性辺と関係を**辺**だと考え，接続関係を表す**連結表現**と，包含関係を表す**領域表現**を次に与える．

連結表現・・・頂点と辺をつないでグラフのように表現．
領域表現・・・辺を包含を表す囲いだと考えて，頂点の中に頂点が含まれるように表現．

has_car(+A, －B)	has_load(+A, －B)
列車 A は貨車 B を持つ	貨車 A は貨物 B を持つ
個体：A, B	個体：A, B
属性：なし	属性：なし
関係：has_car	関係：has_load
load_num(+A, ＃B)	triangle(+A)
貨物 A は B 個である	貨物 A は三角形である
個体：A	個体：A
属性：B(頂点), load_num(辺)	属性：triangle(頂点)
関係：なし	関係：なし

表１:モードから成る対象の例

４．実行例

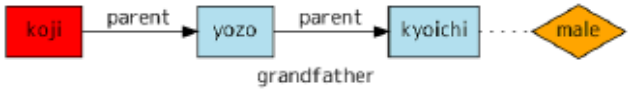

図４:連結表現

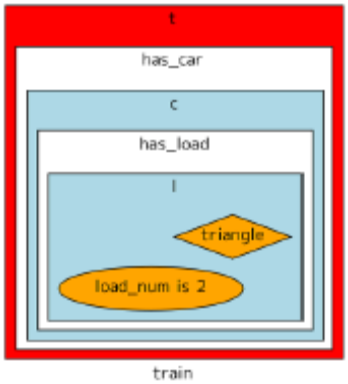

図５:領域表現

当研究室のマイニングシステムに組み込んで，パターンを機械的に変換することができる(適宜頂点や辺の形や色を変えて理解しやすい様にしている(図４，図５))．

５．まとめと今後の課題

本研究では MRDM におけるパターンの視覚化においてパターンを意味的に**個体**，**属性**，**関係**という 3 つの対象に分類することにより接続関係や包含関係を表す**連結表現**や**領域表現**での視覚化が可能となった．今後の課題としては，想定していないパターンの視覚化や，視覚化を用いた新たな規則の発見，実体関連モデルについての検討が挙げられる．

参考文献

[1]吉野 伶,“関係型データマイニングのための統合的環境の構築”，名古屋工業大学平成 20 年度卒業論文，2008.