

属性値の再帰的な評価による分類方法*

5 P-1

原智美, 乾伸雄, 小谷善行, 西村恕彦
 (東京農工大学 工学部 電子情報工学科)

1 はじめに

フレームで表現された知識の特徴継承階層はさまざまな推論に利用される。したがって分類されていないフレームの階層への位置付けは重要である。

同類関係のある二つのフレームは、同じスロット名のスロットを持ち、そのスロット値にも同類関係があると考えられる。そこでスロット値になっているフレームを再帰的に評価する分類方法を考案した。本稿では二つのフレームの同類関係の評価方法と、上位フレームのスロット決定方法について述べ、それを検証する実験と結果を述べる。

2 同類関係の評価

同類関係の評価は、次に述べるフレーム同類度を計算し、あらかじめ定めておいたしきい値と等しいかあるいはそれ以上の場合、同類関係が成り立つと判定する。

フレーム F_A 、 F_B が持つスロットのスロット名の和集合 $SN(A, B)$ を $SN(A, B) = \{N_{ABi} | 1 \leq i \leq s\}$ とする。ただし N はスロット名、 s はスロット名の種類数。また階層関係を表すスロット名は含まない。

まずフレーム F_A と F_B のスロット名 N_{ABx} のスロット同類度 $C_S(A, B, N_{ABx})$ を次のように定義する。ただしフレーム F_A スロット名 N_{ABx} のスロット値の集合 $SV(A, x)$ を $SV(A, x) = \{V_{Ax\alpha} | 0 \leq \alpha \leq t_{Ax}\}$ とする。 V はスロット値、 t_{Ax} はスロット値の種類数を表す。同様にフレーム F_B のスロット値の集合を $SV(B, x) = \{V_{Bx\beta} | 0 \leq \beta \leq t_{Bx}\}$ とする。

- $SV(A, x) = \phi$ あるいは $SV(B, x) = \phi$ のとき

$$C_S(A, B, N_{ABx}) = 0$$

- $SV(A, x) \neq \phi$ かつ $SV(B, x) \neq \phi$ のとき

$SV_C(A, B, x) = SV(A, x) \cap SV(B, x)$ の要素の数を l とする。

$SV_R(A, x) = SV(A, x) - SV(B, x) = \{V_{RAxj} | 0 \leq j \leq m\}$ とする。要素の数は $m = t_{Ax} - l$ 。同様に $SV_R(B, x) = SV(B, x) - SV(A, x) = \{V_{RBxk} | 0 \leq k \leq n\}$ とする。要素の数は $n = t_{Bx} - l$ 。

$m \neq 0$ かつ $n \neq 0$ であれば、 $SV_R(A, x)$ と $SV_R(B, x)$ の組合せについてフレーム同類度を計算し、その平均を $A_{ABx} = (\sum_{j=1}^m \sum_{k=1}^n C(V_{RAxj}, V_{RBxk})) / mn$ とする。
 $m = 0$ あるいは $n = 0$ であれば $A_{ABx} = 0$ 。

スロット同類度 $C_S(A, B, N_{ABx})$ を計算する。

$$C_S(A, B, N_{ABx}) = 0.5 + \frac{2l + mA_{ABx} + nA_{ABx}}{2(2l + m + n)}$$

フレーム F_A 、 F_B のフレーム同類度 $C(A, B)$ を次のように定義する。

$$C(A, B) = \frac{\sum_{i=1}^s C_S(A, B, N_{ABi})}{s}$$

フレーム F_A 、 F_B の両方あるいはいずれかが全くスロットを持たないか階層関係を表すスロットしか持たないとき、あるいは $C(A, B)$ が呼出しの過程で再帰ループとなるときは $C(A, B) = 0$ とする。

3 上位フレームのスロット決定方法

上位フレームのスロット構成は次のようにして決定する。同類度のしきい値を α で表し、フレーム F_A と F_B について、 $C(A, B) \geq \alpha$ のときのスロット名 N_{ABx} のスロット値の集合 $SV(A, x)$ と $SV(B, x)$ について考える。

*A Method of Classification by Recursive Evaluation at Attribute Values,
 Tomomi HARA, Nobuo INUI, Yoshiyuki KOTANI, Hirohiko NISIMURA
 Tokyo University of Agric. and Tech., Dept. of Computer Science

1. $SV_R(A, x)$ と $SV_R(B, x)$ の要素のすべての組合せについて、 $C(V_{RAxj}, V_{RBxk}) \geq \alpha$ の成り立つ組を求め、その上位フレーム名の集合を $F_H(A, B, x)$ とする。ただし、 $C(V_{RAxj}, V_{RBxk}) \geq \alpha$ が成り立ち、上位フレームが存在しない場合には、 V_{RAxj} 、 V_{RBxk} の同類関係を生成し、上位フレームを決定する。
2. 上位フレームのスロット構成として、スロット名 N_{ABx} スロット値 $SV_C(A, B, x)$ のスロットと、スロット名 N_{ABx} スロット値 $F_H(A, B, x)$ のスロットを与える。

同様にして $SN(A, B)$ の要素すべてについておこなうことにより、フレーム F_A と F_B の上位フレームのスロット構成が決定する。

ただし 1 において、再帰呼出しによりループが存在する場合には一時的な仮のフレーム名を与え、呼出しが終わってから決定した上位フレーム名に変換する。

4 実験

同類関係の評価方法、上位フレームの生成方法を検証するために実験をおこなった。

まずランダムな階層を生成する。この階層を比較基準の階層とする。ランダムに生成した階層の上位・下位関係と、下位フレームを持つフレームの半分を削除する。上位・下位関係が削除されたフレーム群を未分類フレーム群として、新たに分類する実験をおこなった。そのとき同類度のしきい値の与え方によって、生成される階層がどのように復元されるか調べる。

5 結果

基準の階層中において成り立つすべての上位・下位関係のうち、生成階層中に復元された上位・下位関係の割合を分類階層包含率とした。

また生成階層中において成り立つすべての上位・下位関係のうち、基準の階層中では成り立たない上位・下位関係の割合を上位・下位関係過生成率とした。

フレーム数 10 のランダムな階層 25 個について実験した。その平均を図 1 に示す。

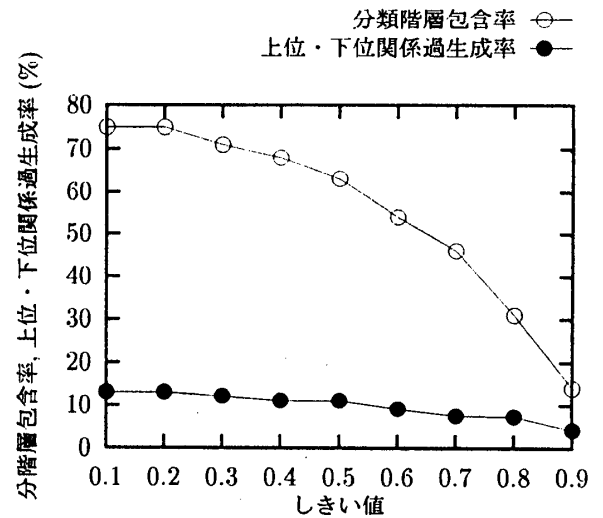


図 1: 分類階層包含率と上位・下位関係過生成率の変化

図 1 から、考案した同類関係の評価方法と上位フレームの生成方法から元の分類階層がある程度復元できることがわかる。

6 おわりに

同類関係の評価方法は各フレームのスロット構成に依存する。したがって有意な分類をおこなうには各フレームのスロットを正確に表現する必要がある。

過生成された上位・下位関係は新たな分類方法を表している可能性もある。

同類関係の判定にはしきい値基準にしておこなっており、ある未分類フレーム群に対して適切なしきい値を定めることが難しいという問題がある。

参考文献

- [1] 椎谷秀一, 瀧口伸雄, 小谷善行: フレーム形式の知識ベースにおける単語概念の学習アルゴリズム, 情報処理学会, AI-87-6 (1993)
- [2] 小泉和彦, 大川剛直, 馬場口登, 手塚慶一: クラスフレーム生成によるフレームの階層化, 情報処理学会, AI-65-7(1989)