

分類観点に沿った概念階層全体の比較によるオントロジー品質向上手法の提案

Refine Method to Improve Ontology Quality based on Comparison of Whole Hierarchies

増田 壮志*¹
Takeshi Masuda

古崎 晃司*¹
Kouji Kozaki

駒谷 和範*¹
Kazunori Komatani

*¹ 大阪大学産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

Abstract: The consistency of classification is one of important characteristics that indicate the quality of ontologies. We have been developing a system that supports to improve the consistency through comparing a slot hierarchy and the other two hierarchies that refer to the slot hierarchy. In previous works, we developed a method that compares separated slot hierarchies every one step, and we confirmed that the proposed method can detect refinement candidates appropriately through user experiments. However, the refinement proposals are limited to local ones because we split up slot hierarchies into each specialization, when we compare the three hierarchies. In this paper, we propose a method that compares whole slot hierarchies to the two other hierarchies. It can support users to decide whether a proposal is appropriate or not by considering whole hierarchy about a certain property.

1. はじめに

オントロジーは現在、情報分野のみならず、生命科学、医療情報や環境問題などの様々な領域で構築されており、知識システム開発の知識基盤として用いられている。しかしながら、オントロジー構築及びその洗練は容易ではない。このような需要を背景として、オントロジー構築方法論に関する研究や、構築支援ツールの開発が多く行われてきた。

オントロジーの洗練段階における構築支援は、文法規則のエラー修正と内容洗練の 2 種類に大別できる。前者は、多くの構築ツールで実装されている入力支援や文法チェック機能、整合性検証機能[Sirin 07, Kopena 03, Decker 98 太田 10]等がある。一方、後者の内容洗練に対しては、オントロジーの分類上の関係の妥当性を強化する OntoClean [Guarino 04] やドメイン知識俯瞰のためのオントロジー探索ツールの利用[M. Ohta 11]などがある。

しかしながら、これらの内容洗練手法では、対象ドメインの専門家が人手でオントロジーから洗練箇所及び方法を見つける必要があり、品質向上のための具体的な洗練箇所及び方法を自動的に提示することはできない。そこで本研究では、オントロジーの品質を向上させるための洗練箇所及び方法を自動的に提示できる構築支援手法の開発を目的としている。

本論文では、筆者らがこれまで提案した上位下位関係を利用した概念階層比較手法及び兄弟関係を利用した概念階層比

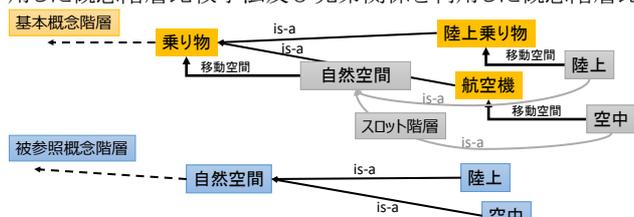


図 1 相似形な is-a 階層

連絡先: 増田 壮志, 大阪大学産業科学研究所 知識科学研究分野, 〒567-0047 大阪府茨木市美穂ケ丘 8-1, Tel:06-6879-8416, e-mail:masuda@ei.sanken.osaka-u.ac.jp

較手法の洗練提案の提示が局所的であるという問題を改善し、分類観点に沿った概念階層全体を提示することによって、ユーザが洗練提案の適用の可否判断を行いやすくする手法を提案する。

2. 概念階層の比較によるオントロジー品質向上手法の現状と課題

筆者らはこれまでも概念階層比較によるオントロジー品質向上手法の開発を進めてきた。本手法は、オントロジー構築における推奨基準の一つである「ある概念の下位概念において、その分類は同じ性質を参照して行われることが望ましい」[溝口 06] に着目し、この指針に沿って構築されたオントロジーに現れる、参照関係にある 3 つの is-a 階層が互いに相似形となるという性質(図 1)を利用する手法である。

ここで参照関係にある 3 つの is-a 階層とは、(1)概念定義の際に各概念の定義内容を表し概念の特殊化に伴って特殊化されるスロットの階層、(2)それらのスロットを持つ概念の階層(基本概念階層)、および、(3)スロットから参照されている概念階層(被参照概念階層)の 3 つの階層のことである。

これまでの手法における概念階層の比較は、スロット階層を上位-下位一段毎の単位に区切り、そのスロット階層を持つ基本概念階層及び参照されている被参照概念階層と比較することで行っていた[増田 17]。本手法の評価実験により、特殊化に伴う概念階層の比較によって、オントロジーの品質向上につながる洗練箇所及び方法を提案できることが確認された。

さらに、その評価実験におけるユーザアンケートから得られた「他の階層を比較しながら洗練提案の選択を行いたい」という要望に対応するために、上位下位関係に加えて、兄弟概念を比較対象とする洗練手法を導入した[増田 16]。

しかしながらこれらの手法は、いずれもスロット階層を区切って比較しているため、提案された洗練方法の適用の可否判断が、局所的になってしまうという問題点があった。

本論文では、提案される洗練方法が局所的なものであるという問題に対処するために、スロット階層を一段ごとに区切らず階層全体として他の 2 階層と比較することで、洗練提案の際にユ

一ザに対して分類観点毎に包括的な情報を提示する手法を提案する。

3. 概念階層全体の比較による洗練支援

3.1 分類観点毎の概念階層比較

図1の is-a 階層を例とすると、スロット階層は、乗り物がどこを移動するかを定義する”移動空間”という分類観点で特殊化されている。このように基本概念階層中の各概念は、スロットによってその詳細な定義が表現されており、概念の特殊化の際にはスロットの特殊化によって上位概念や兄弟概念からの違いが明示化される。そのため、一連のスロットの特殊化階層をその分類観点を示すロール名毎に辿ることによって、分類観点毎にどのように基本概念階層が特殊化されたかが分かる。こうした分類観点は法造上¹で、スロットのロール名として表現される。よって、本提案手法では、同じロール名か特殊化されたロール名を持つ一連のスロット階層を抜き出し、それらのスロットを持つ基本概念階層、および、それらのスロットから参照されている被参照概念階層と比較する。

本手法による洗練箇所検出、および、洗練方法の提案は次のような流れで行われる。

1. ある最上位スロット ST とそのすべての下位スロット SL_m ($m = 0 \sim N$, N : ST の下位スロットの総数)を比較元となるスロット階層とする。
2. 最上位スロット ST 及びその下位スロット SL_m を持つ概念とそれらの下位概念からなる基本概念階層と、スロット階層から参照されている概念階層とその兄弟概念からなる被参照概念とを比較し、非相似形となる箇所を洗練候補箇所として検出する。
3. それらの非相似形な箇所のパターンに応じた、具体的な概念及びスロットの追加提案を行う。

このように、スロット全体と他の 2 つの概念階層を比較することによって、ユーザは提案箇所毎に考慮すべき分類観点に沿った一連の概念階層をみることができ、以前手法[増田 16, 17]よりも利用可能な情報が増え、可否判断が行いやすくなると期待される。また、この 1 から 3 の手順を全ての最上位スロットに対して行うことでオントロジー上の全スロットを網羅できるので、以前手法での提案箇所は全て含まれる。このとき今回の手法によって検出される洗練箇所には、以前手法によって検出される複数

の箇所が含まれることになる。

但し、今回の論文では、洗練提案を基本概念階層とスロット階層に対する概念及びスロットの追加提案に留めることにする。なぜなら、以前に行った兄弟概念階層の比較[増田 16]の場合にも問題となった比較対象の概念の組み合わせ爆発による提案数の急増という問題が、本手法でも同様に起こると考えられるからである。そこで本論文では、問題を切り分けて考えるために、まずは被参照概念階層を固定して考えることとした。なお、今後、被参照概念階層に対する概念追加提案も含めた洗練手法に拡張する予定である。

3.2 検出される洗練箇所の分類

スロット階層を中心に 3 つの is-a 階層を比較した場合、相似かどうかはそれぞれの階層の特殊化の詳細度で決まる。つまり、特殊化の詳細度の比較は、基本概念階層、スロット階層、被参照概念階層中の各概念数及びスロット数を比較することに相当する。この方法で検出される洗練箇所の分類は以下のようになる。

(a): 基本概念階層とスロット階層の比較

(a1): 基本概念階層の概念数 > スロット階層の概念数

(a2): 基本概念階層の概念数 = スロット階層の概念数

(b): 被参照概念階層とスロット階層の比較

(b1): 被参照概念階層の概念数 > スロット階層の概念数

(b2): 被参照概念階層の概念数 = スロット階層の概念数

(c): 基本概念階層と被参照概念階層の比較

(c1): 基本概念階層の概念数 > 被参照概念階層の概念数

(c2): 基本概念階層の概念数 = 被参照概念階層の概念数

(c3): 基本概念階層の概念数 < 被参照概念階層の概念数

(a),(b)にスロット階層の方が詳細な場合が存在しないのは、スロットはそれを持つ概念及び参照先概念が無い場合定義不能であるからである。そのため、基本概念階層および被参照概念階層に含まれる概念数は、常にスロット階層中のスロット数と同数か多くなる。これらの組み合わせは、(a)2種類×(b)2種類×(c)3種類であるので、全 12 組である。

3.3 提案される洗練方法

(1) 洗練箇所と提案される洗練方法の関係

検出された洗練箇所に対して提案される洗練方法は、他の階層を比較した際に欠落していた部分への概念及びスロット追加である。例えば、(a1)であれば、基本概念階層がスロット階層よりも詳細度が大きいので、スロット階層への概念追加が提案される。3.2. で述べた各分類による洗練方法の対応は、表1のように決まる。

これらの方法は、検出箇所の構造の分類の組み合わせとなる。例えば、(a1), (b2), (c3)の組み合わせであれば、基本概念階層への概念追加とスロット追加の両方が提案される。組み合わせを表1に示す。但し、(a2)-(b2)- (c2)の組み合わせは、すべて同じ詳細度であり、各概念階層は相似形であるため洗練箇所として検出しない。また今回は、被参照概念階層を固定しているので、(a2), (b2), (c1)には、提案される洗練方法はない。なぜなら(c1)は被参照概念階層の方が少ないが、今回は被参照概念階層を固定しており、また概念階層へは、新規概念の追加提案のみが行われるため、(c1)に対しては何も提案されず、(a2),(b2)は相似形であるため同様に何も提案されないからである。そのためこの組み合わせも洗練箇所として除外する。そのため、洗練候補箇所として検出すべきは残りの 10 パターンとなる。

表 1 洗練箇所と洗練方法の関係

分類			提案される洗練方法
a	b	c	
1	1	1	スロット追加
1	1	2	スロット追加
1	1	3	基本概念+スロット追加
1	2	1	スロット追加
1	2	2	スロット追加
1	2	3	基本概念+スロット追加
2	1	1	スロット追加
2	1	2	スロット追加
2	1	3	基本概念+スロット追加
2	2	1	洗練箇所として検出しない
2	2	2	洗練箇所として検出しない
2	2	3	基本概念+スロット追加

¹ <http://www.hozo.jp/hozo/>

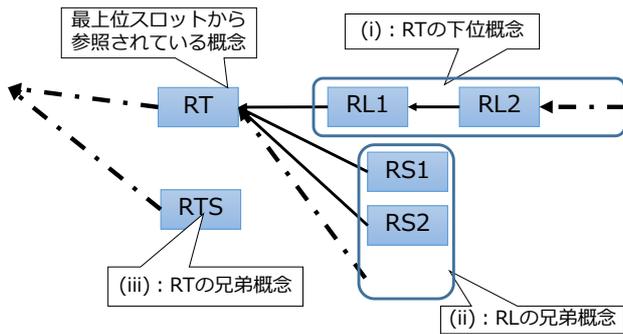


図 2 被参照概念階層の構造

(2) 被参照概念階層の構造による詳細な提案パターン

今回の提案手法では、被参照概念階層を固定するため、具体的な概念及びスロットの追加先、参照する概念の選択は被参照概念階層の構造によって分類できる。図 2 のように最上位スロットから参照されている概念 RT がどのように特殊化されているかで分類可能である。

- (i): RT の下位概念(RL)の有無
- (ii): RT 下位概念の兄弟概念(RS)の有無
- (iii): RT の兄弟概念(RTS)の有無

(i)の場合には、図 3 に示すような概念及びスロット追加が提案される。基本概念階層及びスロット階層よりも被参照概念階層の方が詳細に定義されているため、BL1 の下位に概念を追加し、新たなスロットで RL2 を参照するという提案がなされる。この例では、RL2 が参照されていないが、SL1 によって RL2 が参照されていた場合、BT と BL1 の中間に概念追加が提案されることになる。

(ii)の場合には、図 4 に示すような追加提案がなされる。今回は兄弟概念 RS1 が存在するので、BL1 の兄弟概念を新たに追加し、その概念にスロットを追加し RS1 を参照させるという提案を行う。

(iii)の場合には、図 5 に示すように、BT の兄弟概念の追加が提案される。さらに ST の上位スロットを BT の上位概念 BU(BUpper)に追加し、新規に追加した BT の兄弟概念に下位スロットを追加し、RTS を参照することで ST の兄弟スロットとすることを提案する。

実際は、これら(i) ~ (iii)の提案が個別にされることなく複数混ざった状態で行われることとなる。

また、今回提案した洗練手法は現状では 3.2 節までに述べた洗練箇所の検出までが実装済みである。4 章の評価の際には、出力された洗練箇所に対して、ここまで述べてきたパターンに沿った洗練提案を手動で加えて提示した。

4. 本手法の評価

4.1 目的・概要

ここまで述べてきた分類観点毎の概念階層比較を行う手法の評価を行う。評価対象として用いるオントロジーは、第一著者が本研究を始める前にオントロジー構築技術取得のために構築したレースオントロジーを用いた。このオントロジーの総概念数は 213 個である。評価観点は、提案される洗練箇所の差異と分類観点毎の洗練箇所提案を行うことによる洗練提案の可否判断の違いである。

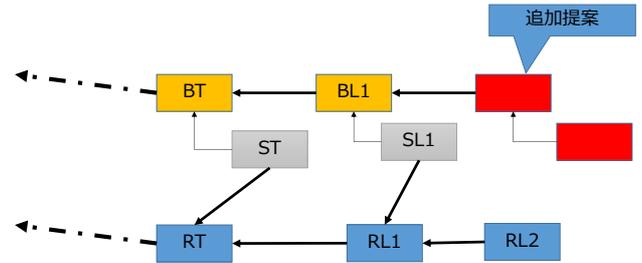


図 3 下位概念がある場合

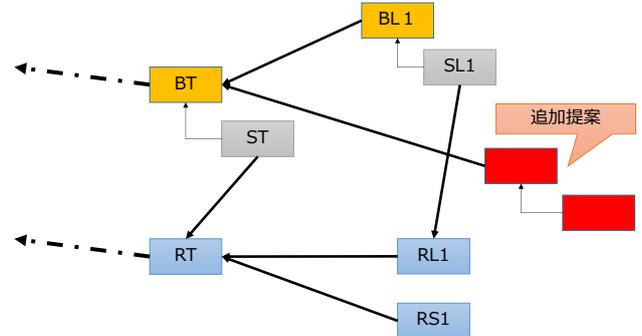


図 4 兄弟概念がある場合

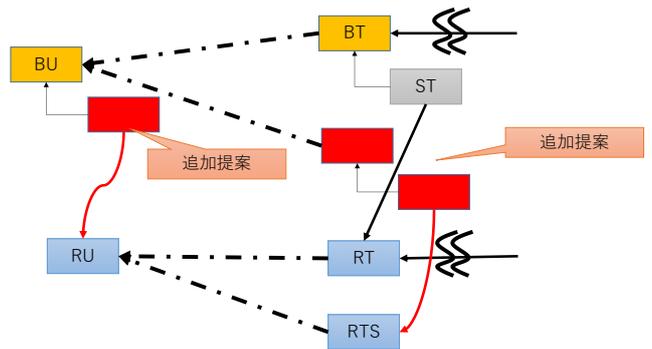


図 5 最上位概念に兄弟概念がある場合

4.2 結果及び考察

(1) 検出された洗練候補箇所の比較結果

今回の手法によって検出された洗練候補箇所数は、72 箇所である。オントロジー中に含まれる最上位スロット 92 個中、先に述べた 10 種類の洗練箇所に対応するものが 72 個であり、それらが洗練提案の対象となる is-a 階層の組である。

今回の手法では、図 6 に示すように一つの洗練箇所、[増田 17]の手法によって検出される複数の洗練箇所が含まれる。図 6 中の赤枠は、以前手法で区切られていたスロット階層を示したものであり、それぞれのスロット階層に伴う概念階層が別々の洗練候補として検出されていた。つまりこの場合以前手法で検出された 3 つの洗練箇所が、本手法では、1 箇所として検出される。これが以前手法のほうが洗練箇所数は多いにもかかわらず、すべて今回の手法で検出された箇所にも含まれ



図 6 以前手法における複数の洗練箇所の検出

る理由である。

[増田 16]の手法では、140 個の洗練箇所が検出されていた。これが 72 の洗練箇所となったのでユーザが見るべき洗練箇所が約 1/2 になったといえる。また、このうち多いものでは 10 以上の洗練箇所が1つの洗練箇所として検出された。

(2) 洗練提案の可否判断の差異

次にスロット階層全体を比較した場合の提案の違いについて述べる。図7に洗練箇所の基本概念階層とスロット階層、図8に洗練箇所の非参照概念階層の例を示す。スペースの関係上、説明に関係する部分のみを抜粋して載せている。この例では、図7中で赤く示されている部分行為スロットを最上位スロットとして持つスロット階層と他の階層を比較している。以前手法では、“運転行為”と“減速”の中間概念“速度調整”に、被参照概念階層中の“シフトチェンジする”を参照した中間スロットを追加するという提案と、“運転行為”と“加速”の中間概念“速度調整”に、被参照概念階層中の“シフトチェンジする”を参照した中間スロットを追加するという2つの提案が別の洗練箇所として提案される。そのためユーザはこれらの提案をそれぞれ別の箇所として提案の可否を判断する必要がある。

一方、本手法では、これらの箇所が一つの箇所として提案されるため、“加速”、“減速”の両方を考慮して判断できる。そのためそれぞれの共通の中間スロットとして“シフトチェンジする”を参照したスロットの追加を考慮することができるようになった。このことは、別々の箇所として提案するよりも、分類観点毎に全ての概念階層を比較する本手法のほうが、一つの洗練提案により多くの包括的な情報を提示することができたと言える。

(3) 考察

スロット階層全体を表示することによって、以前までの手法による洗練提案を包含しながらも、ユーザが判断すべき洗練箇所をまとめて提示することが可能となった。そのことは、以前手法よりも洗練提案の際に多くの情報をユーザに提示することが可能となった。しかしながら、提案される洗練方法数には、工夫の余地がある。3.3.(2)節末で述べたように現状ではまだ具体的な洗練提案を自動的に提示する部分の実装は済んでいないが、兄弟関係を利用した以前手法でも多いもので 1000 を超える提案がなされるため、その手法を包含する今回手法でも同様の問題が予想される。そのため、洗練提案に優先順位をつけるか、具体的な提案を行わず概念階層のみの提示に留めユーザに判断させる等の工夫が必要である。

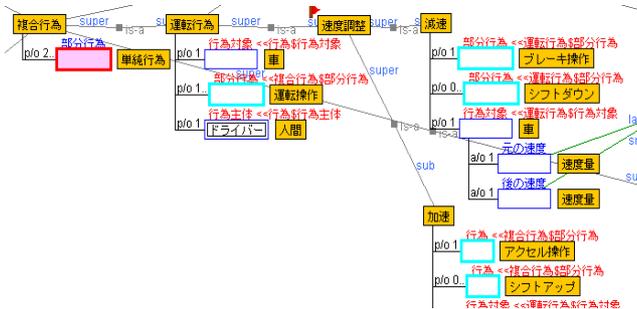


図7 基本概念階層及びスロット階層

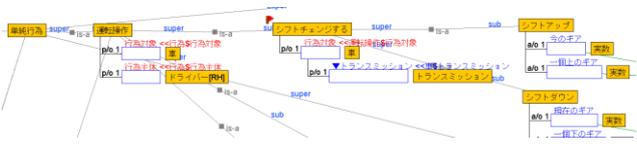


図8 被参照概念階層

また、3.1 節で最上位スロット ST とその下位スロット SL_n を元に各概念階層との比較を行うと述べた。その時複数の最上位スロットが同一の基本概念に存在することがある。今回の評価に用いたレースオントロジーでは、“レース競技”は 7, “車”は 9 と多くなっていた。これらの概念はレースを表現するために中心的概念であった。一方、他の基本概念は 1~3 個程度の最上位スロットのみを持っていた。このことより、オントロジー中で構築者が重要であると考えている概念には最上位スロットが多いという可能性がある。今後、洗練提案の優先順を考えるときの指標の一つとなり得ると考えられる。

5. まとめ

本研究では、スロット階層全体を比較対象とする概念階層比較手法を提案した。以前までの上位下位関係及び兄弟概念関係を利用した洗練手法を発展させ、洗練提案の提示方法が局所的になってしまうという問題の解決を試みた。スロット階層全体を比較対象にした洗練箇所の検出結果から、以前までの手法によって検出されていた洗練箇所及び方法をカバーしつつ、これらの洗練箇所を分類観点毎に統合することによって、ユーザに対してより多くの提案の可否を判断する情報を示すことが可能となった。

今後、本手法によって提案される洗練方法の優先順位について考察し、本手法に基づく洗練支援システムの開発を行う。

謝辞

本研究の一部は、科学研究費補助金・基盤研究(B) 25280081 および基盤研究(A)26240033 の支援による。

参考文献

[Decker 98] Decker, S., Erdmann, M., Fensel, D., and Studer, R.: Ontobroker: Ontology based Access to Distributed and Semi-Structured Information, in Database Semantics: Semantic Issues in Multimedia Systems, pp. 351–369, Kluwer Academic Publisher (1998)

[Guarino 04] Guarino N. and Welty C. An overview of OntoClean. Handbook on ontologies (pp. 151–159). Berlin; Heidelberg: Springer. 2004

[Kopena 03] Kopena, J. B. and Regli, W. C.: DAMLJessKB: A Tool for Reasoning with the Semantic Web, IEEE Intelligent Systems, Vol. 18, pp. 74–77 (2003)

[M. Ohta 11] M. Ohta, K. K. and Mizoguchi, R.: A Quality Assurance Framework for Ontology Construction and Refinement, in 7th Atlantic Web Intelligence Conference (AWIC2011) (2011)

[Sirin 07] Sirin, Evren and Parsia, Bijan and Grau, Bernardo Cuenca and Kalyanpur, Aditya and Katz, Yarden. Pellet: A practical OWL-DL reasoner. Journal of Web Semantics, 5(2):51–53, 2007.

[太田 10] 太田衛, 古崎晃司, 溝口理一郎: オントロジーの学習の現状と動向, 人工知能学会論文誌, Vol. 126, pp. 403-418 (2010)

[増田 17] 増田壮志, 古崎晃司: 概念階層の比較に基づくオントロジー品質向上支援手法, 人工知能学会論文誌, Vol. 32 No. 2, (2017)

[増田 16] 増田壮志, 古崎晃司, 駒谷和範: 兄弟概念の分類観点に着目したオントロジー貫性向上手法の開発, 人工知能学会第 30 回全国大会 (JSAI2016), 2016.

[溝口 06] 溝口理一郎: オントロジー構築入門, オーム社(2006)