

Is-a 階層の相似性を双方向に利用したオントロジー内容洗練方法の考察

Consideration about Ontology Refine Method Using Similarity among Is-a Hierarchies

増田 壮志*^{1,2}
Takeshi Masuda

古崎 晃司*²
Kouji Kozaki

*¹ 大阪大学大学院工学研究科*² 大阪大学産業科学研究所
Graduate School of Engineering Osaka University#1
The Institute of Scientific and Industrial Research(ISIR), Osaka University#2

Quality of ontology is important because it is connected directly with the performance of an application system using the ontology. However ontology refinement to improve its quality needs knowledge and experiments in ontology development. Therefore, ontology refinement task is too difficult especially for beginners in ontology building. In order to solve this problem this article proposes an ontology refinement support system based on similarity among is-a hierarchies and an evaluation of it. The system can support content refinements for ontologies.

1. はじめに

近年、オントロジー工学は、医療情報や機械設計などの様々な領域のオントロジーが構築され、応用システムのモデルや、知識基盤として用いられている。オントロジーの品質はそれを利用する応用システムに直接的に影響を及ぼす重要な要素である。そのためより良質なオントロジーの構築は重要な課題である。

しかしながら、オントロジーの構築には、オントロジーそのものについての知識や経験に加え、オントロジーを構成しようとしているドメインの専門知識が必要である。そのため、オントロジー構築初心者にとって、構築は容易なことではなく、オントロジー構築方法やその支援方法の確立が望まれている。

オントロジー構築支援には、まずオントロジーの大枠の構築を支援する方法がある。オントロジー構築プロセス全般のガイドラインの提案[Noy 01]や、Wikipediaの半構造化情報[森田 10]などの既存の情報を利用した半自動構築などである。一方、オントロジーの品質はそれを用いたシステムの能力に影響を及ぼすため、構築されたオントロジーの品質向上のための洗練支援を行うといった方法も考えられている。本研究では、ある程度構築されたオントロジーに対して、その品質を向上するための洗練作業についての考察を行う。

2. オントロジー洗練支援

オントロジー洗練方法としては、オントロジーの文法規則に関する形式的なエラーの検出、修正をする方法、オントロジーの内容を洗練する方法の2種類がある。オントロジーの形式的なエラー修正に関しては、推論機構の整合性機能を用いた手法が多く提案されている[太田 11, Ohta 11]、また owl 上でよく見られる間違いを列挙し、その幾つかについて修正をするといったものがある[Poveda 12]。一方、オントロジーの内容洗練を対象としたものは少なく、また属人的な手法にとどまっている。これらのアプローチでは、内容評価にドメインの知識やオントロジー構築に関する知識が不可欠であり、オントロジーの内容洗練に実質的に支援しているとは言い難い。

そこで本研究では、ドメインの専門家の知識やオントロジー構築に関する知識の有無に関係なく利用可能な内容洗練手法を提案する。

3. Is-a 階層の相似性を利用したオントロジー洗練

3.1 Is-a 階層の相似性

オントロジー内には、その性質上 is-a 階層が部分的に相似形になっている部分が多く見られる。良いオントロジーを構築する際の指針として「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]がある。例えば、図1の is-a 階層を考える。「乗り物」はその下位で「陸上乗り物」と「航空機」の2つの概念に特殊化されているが、この時、それぞれの概念においてクラス制約として参照している概念はそれぞれ「自然空間」、「陸上」、「空」であり、どちらも移動空間で分類され、スロットも特殊化されている。同様に、「自転車」もクラス制約として参照している概念の is-a 階層に沿って、そのクラス制約が特殊化されている。このように、ある概念を特殊化するときはそのスロットも特殊化される。ここで「クラス制約の特殊化に伴うスロットの階層」を「スロット階層」、「スロットのクラス制約となっている概念の階層」を「被参照概念階層」と以下では呼ぶことにするが、それらは相似形となる。さらにスロットの階層はそのスロットを持つ基本概念階層に沿って形成されるため、オントロジー内では、「基本概念階層」、「スロット階層」、「被参照概念階層」の3つの階層が部分的に相似形となっている箇所が多く見られると言える。

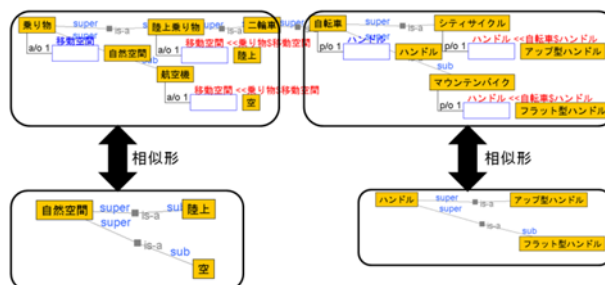


図1 is-a 階層の相似性

3.2 Is-a 階層の相似性を利用したオントロジー洗練支援

前節で述べた3つの階層「基本概念階層」、「スロット階層」、「被参照概念階層」の相似性をオントロジーの内容洗練に対する支援に利用する際には、2つの方向性が考えられる。

- (1). 基本概念階層の洗練を提案する方向. すなわち「被参照概念階層」と「スロット階層」を比較し、それら 2 つの階層が相似形となるような「スロット階層」の洗練を提案し、「基本概念階層」の洗練を行う。
- (2). 被参照概念階層の洗練を提案する方向. すなわち「スロット階層」と「基本概念階層」を比較し、それら 2 つの階層が相似形となるように「スロット階層」を洗練することを示唆し、それに伴う「被参照概念階層」への概念追加を提案する。されに追加された概念を「スロット階層」、「基本概念階層」にそれぞれ反映する。(1)を順方向の洗練支援、(2)を逆方向の洗練支援と呼ぶ。

3.3 順方向の支援

(1) 順方向の洗練支援の概要

順方向の洗練支援の場合、まず被参照概念階層とスロット階層を比較する。次に、どこからも参照されていない概念を支援対象概念とし、その周囲の階層を被参照概念階層として注目する。被参照概念階層には、参照されていない概念が含まれているため、必然的に比較されるスロット階層とは相似形にならない。

順方向の洗練支援の提案の例として、図 2 の is-a 階層を考える。この場合は、どこからも参照されていない概念を支援対象概念とするが、この例では、「シフトチェンジする」が支援対象概念である。そこでその上位概念「運転操作」と下位概念「シフトダウン」を考える。この 2 つの概念は、「運転行為」及び「減速」において同じ「部分行為」スロットで参照されている。そのため「運転行為」と「減速」の中間概念である「速度調整」に「部分行為」スロットを追加し、そのスロットでクラス制約として「シフトチェンジする」を参照することで、被参照概念階層とスロット階層、基本概念階層を相似形にする」という提案ができる。その提案に沿って概念、スロットを追加することで基本概念階層と被参照概念階層の分類の軸、さらにその分類の詳細化の程度を一致させる。

このように、参照されていない概念を支援対象概念としその周囲の概念構造に注目し、それらの階層と相似形となるようにオントロジーの洗練提案を行う。その時洗練提案が考えられる箇所にはいくつかのパターンがあり、その分類パターンを計算機で判別可能な条件で明示化することによって、支援適用候補先を機械的に推定することが可能となる。



図2 順方向の洗練支援の提案例

(2) 順方向の洗練支援の分類と提案される支援方法

順方向の支援パターンを、注目する「被参照概念階層」の概念構造から次の 3 つに分類した。

1. 支援対象概念とその上位概念と下位概念を被参照概念階層とする。
2. 支援対象概念とその上位概念を被参照概念階層とする。

3. 支援対象概念とその下位概念を被参照概念階層とする。
- さらに、洗練支援方法は、支援の適用先となる基本概念階層に依存するため、その構造からさらに次の 2 パターンに分類することができる。

- (ア) 基本概念階層に支援対象概念をクラス制約としたスロットを追加できる概念が無い。
 - (イ) 基本概念階層に支援対象概念をクラス制約としたスロットを追加できる概念がある。
- (ア)(イ)それぞれの場合に考えられる支援方法は以下の様になっている。

(ア)の場合、既存の概念にスロットを作成し、そこで支援対象概念を参照するという洗練方法と、その概念にスロットを追加することが不適切であると考えられるときには、支援対象概念を参照するスロットを持つのに適当な概念を新たに定義して、その概念にスロットを追加する。

(イ)の場合、スロットを追加する概念が存在しないため、新たな概念を定義し、その概念に支援対象概念を参照するスロットを追加する。

以上から順方向の洗練支援は、1~3 と(ア)(イ)の組み合わせから計 6 パターンに分類することができる。

3.4 逆方向の洗練支援

(1) 逆方向の洗練支援の概要

逆方向の場合、まず「スロット階層」と「基本概念階層」を比較する。次にそれらの階層を相似形にするために必要な概念を被参照概念階層に追加し、スロット階層と基本概念階層を相似形にする。例として図 3 と図 4 の is-a 階層を考える。図 3 では、「自動車競技」「フォーミュラーカー」「F1 レース」「F3 レース」「インディ 500 レース」が基本概念階層を構成し、その「使用車両」スロットが特殊化されスロット階層を構成している。しかしこの時、それら 2 つの階層は相似ではない。そこでその制約概念階層である図 4 の is-a 階層に「フォーミュラーカーレース」の「使用車両」スロットのクラス制約となるような新たな中間概念を作成するという提案を行う。逆方向が順方向と異なっている点は、支援適用の際に必ず概念定義が伴うことである。それは、逆方向の洗練支援の場合に、スロット階層と基本概念階層を相似形にするためには、被参照概念階層に新たな概念を定義しなければならないためである。

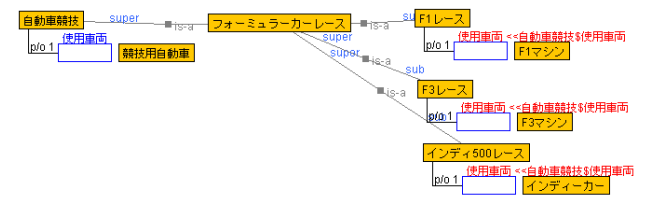


図3 基本概念階層とスロット階層

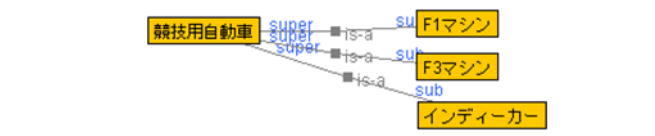


図4 被参照概念階層

(2) 逆方向の洗練支援の分類と提案される支援方法

逆方向の洗練支援の場合では、注目するスロットを持つ基本概念階層の概念構造から次の 3 つに分類分けした。

- (1). スロット 2 つから構成されたスロット階層に注目し、それらを持つ基本概念階層を考慮する。

表 1 提案手法の適用範囲の評価結果

	初心者が構築したオントロジー									熟練者が構築したオントロジー				
	race1	race2	drums	drums2	traffic	traffic2	rail	rails2	hero.1	平均	soccer	vehicle	YAMATO	平均
分類 1 (ア)	1	2	1	1	6	7	2	5	0		6	1	10	
分類 1 (ア)/全支援候補	1%	1%	1%	1%	4%	4%	3%	4%	0%	2%	3%	2%	2%	2%
分類 1 (イ)	2	9	2	2	0	10	3	5	0		0	3	24	
分類 1 (イ)/全支援候補	3%	6%	2%	2%	0%	6%	4%	4%	0%	3%	0%	5%	6%	4%
分類 2 (ア)	32	44	23	23	81	80	27	62	67		68	32	148	
分類 2 (ア)/全支援候補	42%	32%	24%	27%	55%	46%	38%	47%	48%	40%	39%	50%	36%	42%
分類 2 (イ)	22	34	40	30	20	23	14	19	68		39	6	120	
分類 2 (イ)/全支援候補	29%	24%	43%	36%	14%	13%	20%	14%	49%	27%	23%	9%	29%	20%
分類 3 (ア)	3	1	0	0	1	1	0	0	1		0	0	4	
分類 3 (ア)/全支援候補	4%	1%	0%	0%	1%	1%	0%	0%	1%	1%	0%	0%	1%	0%
分類 3 (イ)	17	49	28	28	39	53	25	42	4		60	22	104	
分類 3 (イ)/全支援候補	22%	35%	30%	33%	27%	30%	35%	32%	3%	27%	35%	34%	25%	31%
集計														
概念の全数	133	212	135	135	193	251	112	175	50		261	122	561	
全参照無し概念数	79	88	58	58	87	69	77	97	29		148	75	363	
支援適用先推定可能概念数	34	52	34	34	40	41	29	36	11		55	28	101	
支援適用先候補の総数	77	139	94	84	147	174	71	133	140		173	64	410	
支援対象概念の割合	59%	42%	43%	43%	45%	27%	69%	55%	58%	49%	57%	61%	65%	61%
支援適用先候補提示率	43%	59%	59%	59%	46%	59%	38%	37%	38%	49%	37%	37%	28%	34%

※同一の支援対象概念に対して、複数の支援適用先候補が提案されることがあるため、「支援適用先推定可能概念数 (少なくとも 1 つの支援適用先候補が推定された支援対象概念の数)」と「支援適用先候補の総数」は一致しない

(2). スロット 1 つに注目し、そのスロットを持つ基本概念階層とその上位概念に注目する。

(3). スロット 1 つに注目し、そのスロットを持つ基本概念とその下位概念に注目する。

さらにその支援を適用する先の被参照概念階層を考慮すると、それぞれ次の 2 パターンが考えられる。

(a). 注目しているスロットが参照している被参照概念階層と基本概念階層が相似形となっており、被参照概念階層上にスロットのクラス制約として適切な概念が存在する。

(b). 注目しているスロットが参照している被参照概念階層にスロットのクラス制約として適切な概念が存在しない。

それぞれに提案される支援方法は、

(a) の場合、既存の概念が基本概念階層とスロット階層を相似形にする際のスロットで参照するのに適当な概念ならばそのままスロットのクラス制約とする、またその概念がクラス制約として適当でないならば、被参照概念階層に新たに概念を定義しそれをスロットのクラス制約として参照させる。

(b) の場合、被参照概念階層に新たに概念を定義し、その概念を参照するスロットを基本概念階層とスロット階層が相似形になるように追加する。以上のような洗練方法がそれぞれに対して考えられる。

逆方向の場合に以上の(1)~(3)、(a)(b)の分類の組み合わせから考えられる支援方法は計 6 パターンとなる。

3.5 洗練支援システムの試作

提案した洗練手法にもとづきオントロジーの内容洗練支援システムを設計・試作した。本システムは、3.3 節で述べた順方向のパターンに基づいて洗練支援を適用する箇所の候補と提案する修正内容を「支援適用先候補推定モジュール」、推定された内容に基づき洗練支援案をユーザに提示し、適用する修正案を選択するための「支援適用先候補表示・選択モジュール」、ユーザが選択した修正案を適用し、オントロジーの定義内容を修正する「支援適用モジュール」から構成される。

今回試作したシステムは、3.3 節の順方向の洗練支援に対するものであるが、3.4 節の逆方向の洗練支援においても、同様のモジュールを用いて洗練支援の推定、提示、修正は可能であり今後実装する予定である。

なお、本システムの実装には Java を用いており、オントロジーを処理するためのライブラリとして HozoCore および法造 OAT (Ontology Application Toolkit) を利用している。

4. 洗練手法の評価

4.1 評価方法

3 章で提案した、オントロジーの内容洗練手法の有用性を評価するために、試作した内容洗練支援システムを実際に構築されたオントロジーに適用する評価実験を行った。

評価の対象としたオントロジーは、初心者が初めて構築したもの 5 つと熟練者のアドバイスを受けそれを修正したもの 4 つ、一般に公開しているオントロジー 3 つの計 12 個である。

評価方法は

(1) 提案手法の適用範囲の評価

順方向の場合、参照されていない概念に注目し、その概念を支援対象概念としたので、適用範囲の評価基準を次の 3 つとした。

(ア) 全概念に対する支援対象概念の割合。

(イ) 全支援対象概念数に対する、支援適用先候補を推定することができた概念数の割合。

(ウ) 推定された支援適用先候補の分類ごとの割合

(2) 推定された支援適用先候補の妥当性の評価

初心者が構築したオントロジーのうち 1 つに対してシステムが推定した支援適用先候補について、「提案された修正方法を適用することが、オントロジーの内容洗練として適切かどうか」を実験者が判断し、適切と思われる提案の割合を調べる。

4.2 結果と考察

(1) 提案手法の適用範囲について

支援適用候補提示率を見ると、提案手法の適用範囲は、初心者が構築したオントロジーは 49%、熟練者が構築したものは、34%の支援対象概念に対して洗練方法を提案することができたことが確認できた。(表 1)

次に推定された支援適用先候補の分類ごとの傾向を見ると、支援対象概念とその上位概念と下位概念の両方からなる階層を被参照概念とするより厳しい条件を満たしている分類1(ア)(イ)の割合が他のものよりも低いことがわかる。これらは分類の条件が厳しい分、提案される支援方法の妥当性が高いと考えられる。また分類 3(ア)では、支援対象概念の下位概念がその

	分類1		分類2		分類3		合計
	ア	イ	ア	イ	ア	イ	
提案された支援適用先の数	2	9	44	34	1	49	139
適用が妥当と思われる支援の数	2	5	2	8	0	11	28
適用が妥当な数の割合	100%	56%	5%	24%	0%	22%	20%

表 2 提案された洗練支援の適用妥当性の評価結果

オントロジー内で最上位の概念から参照されていることが分類条件となっているが、最上位の概念は通常ごく少数であるためここで提案される数が少なくなっていると考えられる。

(2) 推定された支援適用先候補の妥当性について

初心者が構築したオントロジーのうち race2 に対して提案された支援適用先候補の妥当性を確認した。その結果 20%の候補については提案された修正を適用することが妥当であると判断された(表 2)。

妥当性を確認する際に、提案された修正案が不適切であると判断された例としては、支援対象概念が“人工物”であるとき、その上位概念“実在物”が“行為”の対象物スロットで参照されている時に、「行為」の下位概念として対象物が“人工物”である“人工物対象行為”の追加する」という提案がある。しかし、この提案を適用することによって追加される“人工物対象行為”は不自然な概念定義であるため不適切であるとされた。このように概念定義に不自然さを感じるために提案が不適切であると判断された例は、概念そのものの抽象度の高いものに多く見られた。トップレベルに近い上位概念を参照した概念定義は、オントロジーの定義内容として誤りではないが、“人工物対象行為”のように日常的には概念化されないような不自然なものが多い。

そこで、支援対象概念からトップレベルに近い概念を除くことで、オントロジーの定義上誤りではないが、不適切であると判断される可能性が高い支援方法を除いた。すると、表 3 の結果で示されるように、提案された修正を適用することが妥当であると判断される割合が 39%に向上した。このことからトップレベルの概念参照を考慮することで、提案される修正案の妥当性を向上させることができることが確認できる。

	分類1		分類2		分類3		合計
	ア	イ	ア	イ	ア	イ	
提案された支援適用先の数	2	5	25	18	0	21	71
適用が妥当と思われる支援の数	2	5	2	8	0	11	28
適用が妥当な数の割合	100%	100%	8%	44%	0%	52%	39%

表 3 提案された洗練支援の適用妥当性の評価結果

(トップレベルの概念を支援対象概念から除いた場合)

4.3 提案手法の有用性についての考察

本手法では、提案された洗練方法を実際の洗練作業に適用するか否かはオントロジー構築者の判断に委ねられる。そのた

め提案の妥当性が 39%であれば、十分有用であると思われる。さらに、適用が不適切と判断した洗練提案についても、オントロジーの定義内容としては誤りで無く、日常的な概念化とのギャップによる不自然さが主な問題であり、オントロジー洗練作業を大きく妨げるものではない。しかし、提案される支援方法の妥当性を構築者が判断する必要があり、不適切であることが多いトップレベル概念の場合、日常的な概念化とのギャップが大きくなるため判断はより困難になると考えられる。また不適切であると考えられる支援方法もオントロジーの定義上は誤りではないことも妥当性の判断を困難にする要因であると考えられる。そのため、提案される支援方法の優先度なども併せて考察する必要がある。また、支援適用先候補の妥当性の評価を、初心者が構築したオントロジーの1つに対してのみしか実施しておらず、今後評価対象とするオントロジーの数を増やし、より詳細な評価と考察を重ねる必要がある。また、この評価実験を行ったのは、順方向の場合のみであり、逆方向の洗練支援に対しても評価実験及び考察を行い、提案される支援方法の優先順位や繰り返し支援を適用した場合に推定される支援適用先候補への影響に関する考察をしていく必要がある。さらに、表1における熟練者の構築したオントロジーにも、支援対象概念は多く存在しているが、支援適用先候補の提示率が低くなっていることも考察する必要がある。

5. まとめ

本研究では、オントロジーの中で部分的に相似形となっている is-a 階層に注目し、その相似性を利用したオントロジーの内容洗練支援方法についての考察及び、評価を行った。本手法の特徴は、限定的ではあるが、オントロジーの内容洗練を機械的な支援によって実現していることである。今後、逆方向の洗練支援に対しても評価、考察を行い、順方向の洗練手法と併せて発展させることでより精度の高い洗練支援の実現を目指す。

謝辞

本研究の一部は科学研究費補助金基盤研究(B)25280081の助成による。

参考文献

- [Noy 01] Natalya F. Noy, Deborah L. McGuinness : A Guide to Creating Your First Ontology , Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI, 2001.
- [森田 10] 森田武史, 山口高平, : オントロジーの学習の現状と動向, 人工知能学会誌 Vol.25(3), 2005
- [Ohta 11] Mamoru Ohtai, et al.: A Quality Assurance Framework for Ontology Construction and Refinement, In Proc. of 7th Atlantic Web Intelligence Conference (AWIC2011), pp.207-216, Fribourg, Switzerland, January 26-28, 2011.
- [太田 11] 太田 衛, 古崎 晃司, 溝口 理一郎: 実践的なオントロジー開発に向けたオントロジー構築・利用環境「法造」の拡張 — 実践編 —, 人工知能学会論文誌, Vol.26 No.2, pp.403-418, 2011
- [Poveda 12] Maria Poveda, Mari Carmen Suarez-Figueroa, Asuncion Gomez-Perez : Validating ontologies with OOPS!, Knowledge Engineering and Knowledge Management, 18th International Conference, (EKAW 2012), pp 267-281, Galway City, Ireland, October 8-12, 2012
- [溝口 06] 溝口理一郎, 古崎晃司, 來村徳信, 笹島宗彦: オントロジー構築入門 pp50-52, オーム社, 2006