

オントロジー内の is-a 階層の相似性に基づいた内容洗練システムの開発

A Development of an Ontology Contents Refinement System based on Similarities among Is-a Hierarchies

増田 壮志*1
Takeshi Masuda

古崎 晃司*2
Koji 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

There are some similar part in ontologies which are built under the guideline that “It is desirable to classify lower concepts according to the same kinds of characteristics with references to them”. Focusing on this feature, we are developing the ontology refine support tool. This tool detects unsimilar parts in ontology, then proposes some refinement methods to the users. They choose these proposal and refine their ontology.

1. はじめに

近年、オントロジー工学は、医療情報等様々な領域で応用システムのモデルや知識基盤として利用されている。その中でオントロジーの品質は、それを利用するシステムに直接的に影響を及ぼす重要な要素である。そのためより良いオントロジーの構築は重要な課題であると考えられている。

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の知識はもとより、構築しようとしているドメインの専門知識も不可欠である。そのため、一般にオントロジー構築は容易なものではなく、その構築方法や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関しては、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を向上させるための洗練支援を行うといった方法も考えられてきており、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支援を目的としている。

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、修正をする方法と、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も存在している[Poveda 12]。

本研究では、オントロジーの内容を洗練することを目的として、修正箇所を検出とそれらに関する洗練方法の提示を計算機に行わせ、その結果をシステムユーザが判断し洗練を行う半自動的な内容洗練システムの開発を行った。

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。これは、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]に従った結果を考察することができる。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。

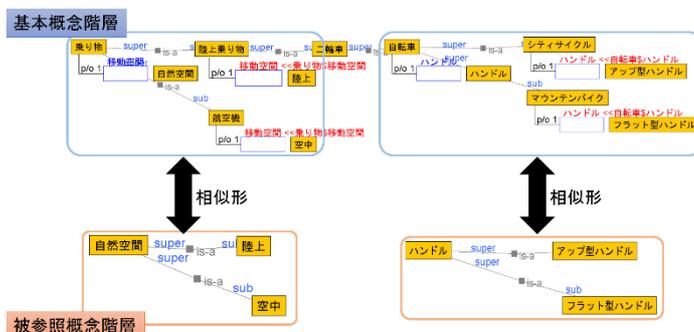


図 1 is-a 階層の相似性

本研究では、これら3つの階層をそれぞれ比較することによって非相似形となっている箇所を洗練支援対象候補と

して検出し、洗練方法の提案を行う。それらをオントロジー構築者に提示することによってオントロジー洗練支援としている。

3.2 スロット階層に着目した、非相似な箇所の検出

(1) 以前の手法との比較

以前から筆者は、is-a 階層の相似性を利用したオントロジー洗練支援システムの開発をおこなっていた[増田 13]が、このシステムは、3.1 節で述べたように3つの概念階層を比較し、修正箇所すなわちオントロジー内で非相似となっている箇所を検出しシステムユーザに洗練方法を提示することでオントロジーの洗練支援としている。これまで、順方向として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」の洗練を行うものと、逆方向として、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洗練手法の2種類について考察を進めていた。

本論文では、それらの手法を発展させた新たな非相似形検出方法として、スロット階層に着目した方法を導入した。スロット階層は、概念を特殊化する際のカテゴリ視点に沿って構築されていくため、基本概念階層や被参照概念階層に着目するよりも的確な検出が可能ではないかと予想される。また、これまでの方法では、基本概念階層または被参照概念階層中のどれか一つの基本概念に着目し、その上位概念と下位概念からなる概念階層と他の階層と比較するという方法を取っていたため、着目した概念の直上の概念と直下の概念のみからなる概念階層になってしまっていた。そのため参照されていない中間概念が多くなっているようなオントロジーに対しては、提案された洗練方法を順次適用しつつ、繰り返し対象としているオントロジーにシステムを適用し再度提案された洗練支援方法を適用するという手段を取らなければならなかった。一方、今回の手法ではスロットの階層に着目するため、基本概念階層、被参照概念階層に複数の中間概念が存在する場合であっても、繰り返し適用していく事無く、検出することが可能となった。

(2) スロット階層に着目した、非相似形の検出の概要

スロット階層に着目して、対象のオントロジーから非相似形となっている箇所を検出するためには、まず、あるスロット階層に着目し、それらのスロットを持つ基本概念階層とそれらのスロットのクラス制約となっている被参照概念階層とを比較し、スロット階層には無い概念が、基本概念階層または被参照概念階層に存在した場合、それぞれ3つの概念階層が非相似形であるため、その箇所を洗練候補箇所として考えられる修正方法と共にユーザに提示する。

例として図2の is-a 階層を示す。いま着目しているスロット階層は、「レース競技」における「自動車」スロットとその下位スロットからなる階層である。いまそれらのスロットを持つ概念からなる基本概念階層は、着目しているスロット階層と相似形である。しかし「自動車」スロットのクラス制約である「レース車両」、「F1マシン」を含む被参照概念

階層は、それらの概念間に「フォーミュラーカー」という中間概念が存在し、スロット階層と相似形とはなっていない。そういった箇所が検出された場合には、それぞれの階層が相似形となるような修正方法を提示する。図2の例の場合であれば、基本概念階層の「レース競技」と「F1競技」の間に「フォーミュラーカー競技」を加え、さらに「フォーミュラーカー競技」に於いて、「自動車」スロットを加えそのスロットのクラス制約を「フォーミュラーカー」とする、といった修正案の提示を行う。

このようにあるスロット階層に着目し非相似形となっている箇所を検出し、それらに対して修正案を提示することによってユーザに基本概念階層、スロット階層、被参照概念階層の分類の詳細化の程度を一致させるような修正を促す。

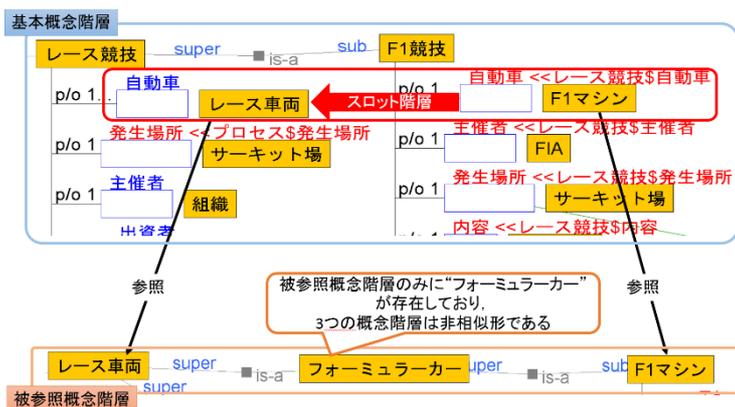


図2 スロットに着目した非相似形箇所検出例

(3) 検出箇所の分類

スロット階層と他の階層を比較する際、考えられるスロット階層の構造にはいくつかの種類があるが、本論文では、着目するスロット階層を次の3種類とした。(a) is-a 階層を構成している2つのスロットからなるスロット階層、(b) 最上位のスロットのみ、(c) 最下位のスロットのみ、の3つである。

また、検出された部分の基本概念階層、被参照概念階層のそれぞれの中間概念の有無によって適用可能な洗練方法が異なるので、

(a) の2つのスロットからなるスロット階層に着目した場合、それぞれの中間概念数の関係から以下に示す3つの方針により分類した。

- 基本概念階層での中間概念の有無
- 被参照概念階層での中間概念の有無
- 基本概念階層と被参照概念階層での中間概念の数の差

それらの観点で分類した結果、

1. 基本概念階層と被参照概念階層のどちらにも中間概念が無い
2. 基本概念階層と被参照概念階層の両方に中間概念が存在し、その数が等しい
3. 基本概念階層と被参照概念階層の両方に中間概念が存在するが、基本概念階層の方が多い
4. 基本概念階層と被参照概念階層の両方に中間概念が存在するが、被参照概念階層の方が多い
5. 基本概念階層のみに中間概念が存在する
6. 被参照概念階層のみに中間概念が存在する

以上の 6 パターンとなった。ただし、1 のどちらにも中間概念が存在しないパターンは、完全に相似形となっているため洗練候補としては提示しない。

次に(b)の最上位のスロットのみに着目した場合、

- そのスロットを持つ基本概念の上位概念の有無
- そのスロットを参照している被参照概念の上位概念の有無

で分類すると、

1. 基本概念階層、被参照概念階層の両方に上位概念が無い
2. 基本概念階層のみに上位概念が有る
3. 被参照概念階層のみに上位概念が有る
4. 基本概念階層、被参照概念階層の両方に上位概念が有る

の 4 種類となった。この場合の 1 も相似形になっているため除外する。

最後に(c)の最下位のスロットのみに着目した場合、(b)の時と同様にして

- そのスロットを持つ基本概念の下位概念の有無
- そのスロットを参照している被参照概念の下位概念の有無

で分類すると、

1. 基本概念階層、被参照概念階層の両方に下位概念が無い
2. 基本概念階層のみに下位概念が有る
3. 被参照概念階層のみに下位概念が有る
4. 基本概念階層、被参照概念階層の両方に下位概念が有る

の 4 種類となった。この場合の 1 も相似形になっているため除外する。

以上より、(a), (b), (c)を併せて、相似形になっているものを除いた計 11 種類となった。

3.3 検出された箇所に対する洗練方法

3.2(3)で述べたように、検出された洗練支援候補に対して、それぞれの概念階層の構造によって適用可能な修正方法が異なる。

(1) 既存の概念を参照させる。

被参照概念階層中の概念を対応する基本概念階層で参照させるという方法である。既存の概念を参照させるためには、被参照概念階層の中間概念が存在することが必要である。また、基本概念階層に参照させたい概念が存在しない場合には基本概念階層に新たな概念を追加し、そこで参照させる必要がある。すなわち 3.3 節の分類では、(a)-2, 3, 4, 6, (b)-3,4, (c)-3,4 の場合に適用可能である。

(2) 新たな概念を定義し、それを参照させる。

この修正パターンは、被参照概念階層で新たに概念を定義し、その概念を基本概念階層に於いて参照させるという修正方法となる。新たに概念を追加するため、3.3 節の全分類に対して適用可能である。ただし、基本概念階層の中間概念の方が多いパターンでは、その中間概念にそって被参照概念階層に概念追加をすれば良いため、他よりも適用が容易であると予想される。

(3) スロット又は概念のレベルを変える

例として図 3 のような場合を考える。この例では、“行為”の“行為対象”スロットと“信号を点灯させる”の“行為対象”

スロットからなるスロット階層に着目している。それぞれのスロットは、“実在物”と“信号”をクラス制約としている。この時、前述の(1), (2)のような洗練方法の提案も可能ではあるが、この場合、“信号を点灯させる”の上位概念である“信号操作”に於いて“行為対象”として“信号”を参照させるという方法も考えられる。この修正方法は、基本概念階層に中間概念が存在するときのみ適用可能である。

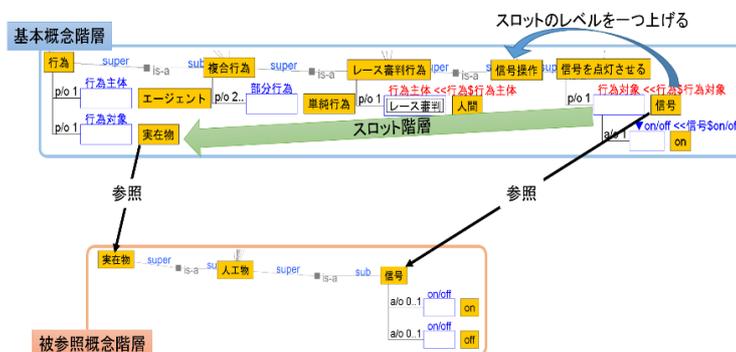


図 3 スロットのレベルを変えるような例

以上の(1)~(3)の修正方法をそれぞれ適用可能な修正先候補に対して提示する。

3.4 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 章で述べたスロット階層とそれに伴う基本概念階層、被参照概念階層に基づいて洗練支援を適用する箇所を検出し、提示する「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。ただし、現在本システムは開発途中であり、現状では 3.2(3)の(a)の 2 つのスロットからなるスロット階層に着目した場合に限定されたものとなっている。

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、スロットに着目した検出方法の有用性を評価するために、3.4 節で述べた内容洗練支援システムを実際にオントロジーに適用し、評価実験を行った。

評価の観点は、以下の 2 点である

[1] 各オントロジーに対する洗練候補検出数について
複数のオントロジーに対して、本システムを適用し、分類毎の数の傾向やオントロジーの規模に対してどの程度数の提案がなされるかを調べる。

評価の対象としたオントロジーは、オントロジー構築初心者が構築した 4 つ(race, boat, drums, dance)と一般に公開しているオントロジー 1 つ(vehicle)の計 5 つである。

[2] 洗練提案の妥当性

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) 各オントロジーに対する検出数について

結果は、表 1 のようになった。各オントロジーに対する洗練候補の数は、今回は 2 つのslotからなるslot階層のみに着目した場合に限っていたこともあり、以前の手法では、オントロジーの規模にも依るが 100 から 400 程度の提案数であったのに対し、2 桁前半に収まる程度の数となった。

また、分類毎の割合に関して見てみると分類 3 と分類 5 は、他に比べ少し高い割合となっている印象を受ける。この 2 つの分類に共通していることは、基本概念階層の方が中間概念の数が多いうことである。基本概念階層は、必要に応じて他の分類の観点での中間概念を挟みながら特殊化していくことを考えると、被参照概念に比べて中間概念が多くなっているのは自然なのではないかと考えられる。

さらに、分類1は、もともと部分的に相似形となっている箇所であるが、全検出数の約 25%~50%を占めている。すなわち 2 つのslotから構成されているslot階層のうち、少なくとも 4 分の 1 程度のものは相似形となっている。このことから本研究で着目したオントロジー内での is-a 階層の部分的な相似性は、実際のオントロジーに於いても現れている性質であることがわかる。

	race	boat	drums	dance	vehicle
分類1	14	20	40	4	5
分類2	3	4	20	1	4
分類3	11	6	17	0	4
分類4	6	0	2	2	1
分類5	9	1	4	5	3
分類6	3	7	3	1	3
Total	46	38	86	15	20
Total (分類1を除く)	32	18	46	11	15
全概念数	212	184	136	211	122

表 1 各オントロジーに対する検出結果

(2) 検出された洗練提案の妥当性について

初心者が構築したオントロジーのうち race2 に対しては、提案された洗練方法についての妥当性を確認した。結果は、表 2 の通りである。表 2 に於いて分類1が書かれていないが、分類1は元々相似形である箇所であり、修正提案を行わないためである。

その結果、44%の候補については提示された洗練方法を適用することが妥当であると判断された。提案された修正案が不適切であると判断されたのは、ほとんどが以前の手法の時と同様に“人工物”や“メタ行為”等の抽象レベルの高い概念での参照やそのような概念の参照が提案されるような場合であった。

また、今回の手法では、2 slot間階層に着目しており、基本概念階層に複数の中間概念が存在している場合も 1 回のツール適用で考慮することも可能であるが、そういった場合に、それらの中間概念は、既にそのslot以外の分類観点で特殊化されており、提案が不適切であると判断されることがあった。

今後、トップレベル概念の参照や他の分類観点を考慮することを考えているが、そうすることでより提案の妥当性を向上させることができると考えている。

分類	2	3	4	5	6	Total
洗練提案箇所数	9	7	3	2	11	32
妥当であると判断された数	3	3	0	1	7	14
提案が妥当であった割合	33%	43%	0%	50%	64%	44%

表 2 洗練提案の妥当性

5. まとめと今後の展望

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 階層の特徴を利用し、slotの is-a 階層に着目することで洗練支援箇所の候補を検出するシステムを開発し、その評価及び考察を行った。以前のある基本概念に着目する方法に比べ効率的に修正箇所を検出することが出来た。本手法の特徴として、限定的ではあるものの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ところにある。

今後の展望としては、今回未実装であった部分の考察を進めると共にトップレベル概念や分類の観点を混在を考慮することで洗練支援候補の検出の精度を向上させたい。また、現状用いている縦方向のslot階層だけではなく、兄弟slotも考慮し、これまでのものと組み合わせ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 「オントロジーの多次的視点管理に基づく領域横断型セマンティックデータの知的探索」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 Ohta, 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
- [増田 13] 増田壮志, 古崎晃司 : 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, 第 27 回人工知能学会全国大会 311-5, 2013