

モノとコトに基づいた知識表現モデル

A Model for Knowledge Representation Based on the Concepts of Things and Events

山田隆弘*1

Takahiro Yamada

*1 宇宙航空研究開発機構 宇宙科学研究所

Japan Aerospace Exploration Agency/Institute of Space and Astronautical Science

To enable automatic processing of human knowledge, various technologies are being studied, such as Knowledge Graph and Linked Data. Several other technologies were studied in the past for representing human knowledge, such as the Production System, Objects, and Frames. These technologies are used to represent knowledge independently of the semantic contents of the knowledge. In order to enable sharing of knowledge, however, these technologies are not sufficient because there can be multiple ways of representing the same knowledge with these technologies. To integrate knowledge bases, each developed by a different group, there must be a guideline or standard that specifies how to represent knowledge based on the semantic contents of the knowledge. This paper proposes a unified model for representing contents of knowledge based on the concepts of things and events. This model combines the object-oriented modeling method with Neo-Davidsonian event semantics.

1. はじめに

人間の持っている知識を機械的に処理できるようにするための技術として Knowledge Graph [Wang 14]や Linked Data [Heath 11]等が注目を集めている。また、以前から知識を形式的に表現する方法として論理式、Production System、オブジェクト、フレーム等の方式が注目されてきた[Brachman 04]。これらの方法は、様々な知識を統一的に表現するための技術として開発されたものであり、知識の内容には依存しない。

知識の内容に依存せずに知識を表現する技術は、知識を機械的に処理できるようにするための基礎として必要な技術である。しかし、それだけでは知識の共有のためには不十分である。なぜならば、一つの(あるいは同様の)知識を Knowledge Graph (あるいは論理式)として表現する方法には、複数のものがあり得るからである。まず、Knowledge Graphなどで使うべき語彙が必ずしも統一されていないという問題がある。世の中には schema.org [schema.org]のような標準語彙集が存在しているが、それが使われない場合もあり得るし、そのような標準語彙集が存在しない分野も多い。さらに、複雑な知識をどのようにグラフ化するのかという基本的な考え方が統一されていない。

一つの知識を複数の異なる方法で表現できるということは、作成元の異なる Knowledge Graph を組み合わせて一つの大きな Knowledge Graph とするような場合、必ずしもうまく組み合わせられないことを意味する。このような状況では、知識の共有や再利用は達成されないであろう。知識の共有を進めるためには、知識の内容に応じて知識をどのように表現するかという基準を作り、同一の基準を適用して様々な知識をグラフ化する必要がある。そこで、本稿では、そのための基礎として、知識内容を表現するためのモデルを提案する。

本稿では、知識内容の基本的な構成要素はモノ(entity)とコト(event)であると考えられる。人間の知識にモノとコトの知識が含まれることは、多くの研究で言及されているが、モノとコトの双方を含めた知識の全体像を表現する方法に関する研究は、ほとんど存在していない。そこで、本稿では、モノとコトの知識を一体的に

表現する方法としてのモデルを提案する。ここで提案するモノの表現方法は、オブジェクト指向モデリング[Bလာ 2005]に基づいている。また、コトの表現方法は、Neo-Davidsonian の事象意味論(event semantics)と呼ばれている方法[Parsons 90]に基づいている。本稿の試みは、この両者を組み合わせることによって知識内容を一体的に表現するためのモデルを制定しようというものである。

なお、本稿は、著者の談話理解に関する研究[山田 18]の中の知識構造に関する部分をさらに発展させたものである。

2. モノに関する知識

2.1 モノ

モノ(entity)とは、ある世界に存在している何かである。ここでいう世界とは、現実の世界でも仮想的な世界でも良い。知識には何かの計画に関する知識も存在する。例えば、鈴木さんが新たに家を建てることを計画している場合、その家は現実には存在していないが、鈴木さんはその家に関する知識を保有しているはずである。このような場合の家もモノであると考えられる。

モノは、一つあるいは複数の属性(attribute)を有している。例えば、家の場合、床面積、外壁の色、価格などを属性として有している。それぞれの属性は、値を有する。例えば、床面積という属性は 120 平米という値を有し、外壁の色という属性は黒という値を有する。

あるモノについて、そのモノの属性と属性値との対を状態(state)と呼ぶことにする。例えば、ある家の「外壁の色が黒である」という言説は、その家の状態の一つを表したものである。

モノは、他のモノとの間に関係(relation)を有することがある。例えば、人が家を所有する場合、人というモノと家というモノの間に「有する」という関係が存在していると考えられる。また、ある家がいくつかの部屋を有している場合、家というモノと部屋というモノの間に「有する」という関係が存在していると考えられる。

ここで定義した概念(すなわち、モノ、関係、属性)は、オブジェクト指向モデリング[Bလာ 2005]で使われているものと同様のものであり、Unified Modeling Language (UML) [Booch 05]の用語に従えば、上記の 3 つの概念は、それぞれ、オブジェクト(object)、リンク(link)、アトリビュート(attribute)に相当する。

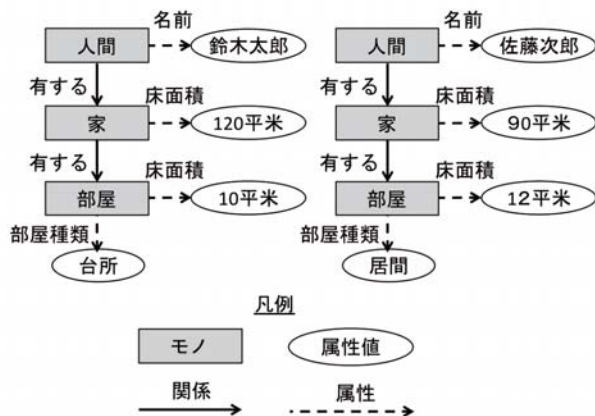


図1 モノを図式的に表した例

以上で定義した概念の例を図式的に表したものが図1である。この図では、モノ、関係、属性、属性値のそれぞれが凡例に示した記号を用いて示されている。これらの概念はUMLのオブジェクト図を用いて表すことも可能である。図1では、鈴木太郎の家も佐藤次郎の家も一つしか部屋を有していないように描かれているが、図を見やすくするために他の部屋は省略されている。また、人間も家も部屋も、この図に示されていない属性も有しているのだが、それらも省略されている。

モノに関する基本的な知識は、上で述べた方法で表現できるが、属性値については、さらに付加的な情報が必要になることがある。それらは、時制(tense)と法性(modality)である。時制については、「太郎の家の外壁は、以前は黒かったが、今は赤い」という知識があり得る。この場合は、個々の属性値(黒や赤)にさらに時間という属性とその属性値(昔、2017年以前、今、2018年5月以降等)を付加する必要がある。法性については、「太郎の家の外壁は、黒いはずである」という知識があり得る。この場合も、属性値(黒)にさらに法性という属性とその属性値(可能性、必然性等)を付加する必要がある。

2.2 部類(モノの概念)

モノには、似たようなモノが存在する。例えば、鈴木太郎の家と佐藤次郎の家は、様々な共通点を有している。いろいろの家は、共通の属性を使って表すことができるが、それぞれの家は、異なる属性値を有している。例えば、鈴木太郎の家の床面積は120平米であるが、佐藤次郎の家の床面積は90平米である。また、どのような部屋を有しているかは、それぞれの家によって異なっているが、どの家も何らかの部屋を有している。

このように共通の属性と関係を有しているモノの集合を部類(class)と呼ぶことにする。あらゆるモノは、何らかの部類に属している。場合によっては、一つのモノが複数の部類に属していることもある。部類は、オブジェクト指向モデリングにおけるクラス(class)に相当する。図1に描かれている「人間」は、実際には、人間という部類に属するモノとしての人間を表している。同図の「家」と「部屋」も同様である。部類は、UMLにおけるクラス(class)に対応する。

部類と部類の間にも関係があり得る。例えば、モノとしての家はモノとしての部屋を有することができる。このような事実は、部類としての家と部類としての部屋の間関係として定義できる。このような部類と部類の間関係は、UMLではアソシエーション(association)と呼ばれている。

図1の表記法に基づいて家と部屋という二つの部類の定義を表したものが図2である。図2に描かれている「家」は、モノと

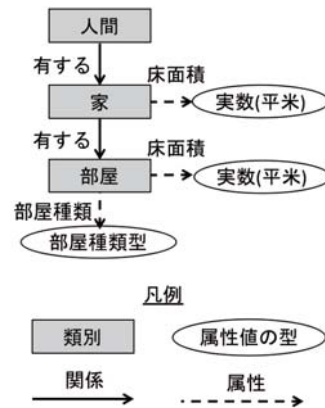


図2 部類(モノの概念)を図式的に表した例

しての家ではなく、部類としての家である。図2では、(1)部類、(2)一つあるいは複数の部類に属するモノの間に成立し得る関係、(3)各々の部類に属するモノが有すべき属性と属性値のデータ型が示されている。これと同様な情報は、UMLのクラス図によっても表現できる。

第2.1節で述べたモノに関する知識は、個々のモノに関する個別的な知識であり、それをここでは個別知識と呼ぶことにする。一方、本2.2節で述べた部類に関する知識は、概念レベルの知識であり、これをここでは概念知識と呼ぶことにする。

3. コトに関する知識

3.1 コト

コト(event)とは、ある世界で生じた何かである。ここでも、世界とは、現実の世界でも仮想的な世界でも良いものとする。

コトは、一つあるいは複数の参加者(participant)を有している。各々の参加者は、モノである。例えば、「鈴木太郎が2018年3月14日に家を買う」は一つのコトである。この場合、鈴木太郎が動作主という参加者であり、家が対象という参加者である。また、コトも属性を有している。「鈴木太郎が2018年3月14日に家を買う」の場合、買った時期は属性であり、その値は2018年3月14日である。また、「鈴木太郎が山本三郎に2018年6月から2020年5月まで月額5万円で家を貸す」もコトである。この場合の参加者は、鈴木太郎と山本三郎と家である。「2018年6月から2020年5月まで」は期間という属性の値であり、「月額5万円」は対価という属性の値である。

ここで定義した概念(すなわち、コト、参加者、属性)は、Neo-Davidsonian 事象意味論(event semantics) [Parsons 90]で使われている概念と類似している。

以上で定義した概念の例を図式的に表したものが図3である。この図では、コト、参加者、属性のそれぞれが凡例に示した記号を用いて示されている。

コトに関する基本的な知識は、上で述べた方法で表現できるが、さらに付加的な情報が必要になることがある。それらは、相(aspect)と法性(modality)である。相については、「太郎は走っていた」という知識があり得て、この場合は、「走る」というコトに相という属性と進行中という属性値を付加する必要がある。法性については、「太郎は走るだろう」という知識があり得る。この場合も、「走る」というコトにさらに法性という属性とその属性値(可能性、義務等)を付加する必要がある。

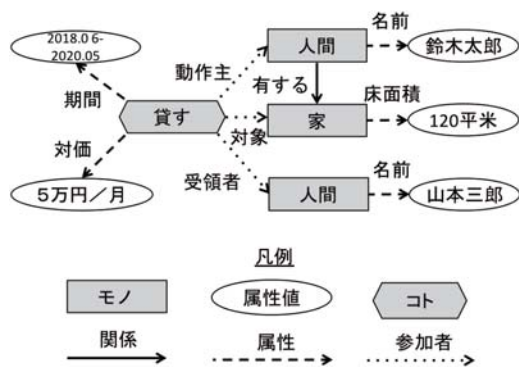


図3 コトを図式的に表した例

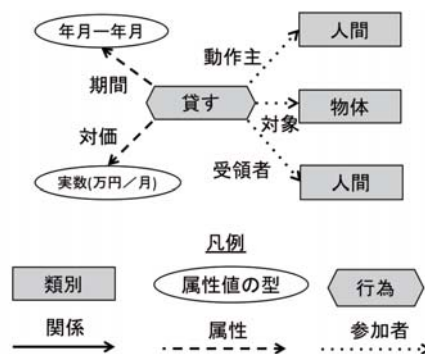


図4 行為(コトの概念)を図式的に表した例

3.2 行為(コトの概念)

コトには、似たようなコトが存在する。例えば、「鈴木太郎が山本三郎に家を貸す」というコトと「佐藤次郎が太田四郎に自動車を貸す」というコトは、様々な共通点を有している。さまざまな「貸す」というコトは、共通の参加者の種類(すなわち、動作主、受領者、対象)を有しているが、それぞれのコトにおける具体的な参加者は、コト毎に異なっている。例えば、鈴木太郎が家を貸す場合は、動作主は鈴木太郎、受領者は山本三郎、対象は家であり、佐藤次郎が自動車を貸す場合は、動作主は佐藤次郎、受領者は太田四郎、対象は自動車である。また、それぞれの「貸す」コトは、共通の属性も有している。

このように共通の参加者と属性を有しているコトの集合を行為(action)と呼ぶことにする。あらゆるコトは、何らかの行為に属している。行為は、Neo-Davidsonian 事象意味論における行為文(action sentence)の述語(predicate)に相当する。図3に描かれている「貸す」は、実際には、「貸す」という行為に属するコトとしての「貸す」を表している。

図3の表記法に基づいて「貸す」という行為の定義を表したものが図4である。図4に描かれている「貸す」は、コトとしての「貸す」ではなく、行為としての「貸す」である。図4では、(1) 行為、(2) 各々の行為に属するコトが有すべき参加者とその部類、(3) 各々の行為に属するコトが有すべき属性と属性値のデータ型が示されている。

第3.1節で述べたコトに関する知識は、個々のコトに関する個別的な知識であり、個別知識に分類される。一方、本3.2節で述べた行為に関する知識は、概念レベルの知識であり、概念知識に分類される。

4. 関係と規則に関する知識

4.1 個別知識としての関係

モノとモノとの間の関係については、第2.1節に述べたように「鈴木太郎が120平米の家を有する」などのものがある。

それ以外にも、個別知識の間には様々な関係が成立する。これらの大部分は、属性値と属性値の間、属性値とコトの間、あるいは、コトとコトの間に成立する。これらの関係は、談話分析の分野で談話関係(discourse relation) [Pitler 08]や修辞関係(rhetorical relation) [Asher 03]と呼ばれているものにほぼ対応する。[Pitler 08]に基づいて幾つかの例を以下に示す。

- 彼は、一日中テニスをしていたので、非常に疲れている。(因果関係)

- 彼は、非常に強くはないが、驚くほど速く走ることができる。(対比)
- 我々は、午後にお茶を飲み、その後、レストランに晚餐会に行った。(時間順序)

これらの関係を図式的に表すには、該当する属性値やコトの間に談話関係を表すリンクを張れば良い。

4.2 概念知識としての規則

モノとモノとの間に関係が存在するかどうかを規定するための概念レベルの知識が規則(rule)である。規則は、複数の部類(あるいは属性値)の間、複数の行為の間、複数の部類と行為の間に成立する。規則には様々なものがあるが、例としては、論理的因果関係(例えば、Xを買うとXを所有するに至る)、物理法則(例えば、水の温度が摂氏0度になると氷に変化する)等がある。また、Schank [Schank 77]のスキーマ(schema)のような社会的な通例に基づく順序関係(例えば、メニューを見てから注文する)も含まれる。規則の中には、非単調論理(例えば、食べると通常は満腹になる)も含める必要がある。

これらの規則を一般的に表現する方法については、今後の検討課題としたい。

5. 知識の全体像

第2節から第4節までに述べてきた知識の全体を一つの表にまとめると表1のようになる。

表1 知識の全体像

	個別知識	概念知識
モノ	<ul style="list-style-type: none"> • モノ • 属性(属性値) • 関係 	<ul style="list-style-type: none"> • 部類 • 属性(属性値の型) • 関係(どの部類間に成立するか)
コト	<ul style="list-style-type: none"> • コト • 参加者 • 属性(属性値) 	<ul style="list-style-type: none"> • 行為 • 参加者とその部類 • 属性(属性値の型)
関係と規則	<ul style="list-style-type: none"> • 談話関係 	<ul style="list-style-type: none"> • 論理的因果関係 • 物理法則 • 社会的通例 • その他

6. 本モデルの意義

本稿では、知識を表現するための一般的な概念とともに、その概念を用いて知識を図式的に表現する方法を提示した。図1

から図 4 までに示したグラフは、そのまま RDF を用いて Knowledge Graph あるいは Linked Data として表現することができる。また、既存の Knowledge Graph や Linked Data も図 1 から図 4 までに示したグラフと同様な構造を持つものが多い。

しかし、単に RDF で書けるというだけでは、知識の共有は困難である。知識の共有を目指すのであれば、第 1 節で述べたように、語彙の統一とともに知識内容を表現するための方法の統一が必要である。本稿で提示したモデルは、知識内容を表現するための統一的な方法を提案するものであり、知識共有を実現するための重要な一つのステップとなる。

本稿では、図 1 から図 4 までに示したように、モノやコトを複数のグラフの集合として表現する方法を示したが、モノやコトを人間に分かりやすい形で扱うためには、モノやコトを一つのまとまった存在(例えば、フレーム)として表現する方が望ましいだろう。フレームとして表現したモノやコトは、グラフにも容易に変換できる。人間に分かりやすい形でモノやコトを表現する方法の開発は今後の課題である。

7. おわりに

本稿では、モノとコトの概念に基づいて知識を一般的に表現するための一つのモデルを提示した。これは、従来から存在しているオブジェクト指向モデリングと Neo-Davidsonian の事象意味論とを一体化させたものである。しかし、この研究はまだ初期段階にあり、今後さらに精密化する必要がある。特に、第 4.2 節で述べたように、規則の一般的な表現方法は今後に残された大きな課題である。また、第 6 節で述べたように、モノやコトを人間に分かりやすい形で表現する方法の開発も今後の課題である。

本稿の研究は、そもそもは談話の意味を記述するための研究の一部として始めたものであるが[山田 18]、この研究の適用範囲はそれだけではない。再利用可能な Knowledge Graph や Linked Data を構築するためには、知識内容に基づいた知識の表現方法に関する統一的な基準が必要であり、本稿の研究は、そのような知識の表現方法の基準作成のための基礎としても使用されるべきものである。

参考文献

- [Asher 03] Asher, N., Lascarides, A.: *Logics of Conversation*, Cambridge University Press (2003).
- [Blaha 05] Blaha, M., Rumbaugh, J.: *Object-Oriented Modeling and Design with UML*, Second Edition, Prentice Hall, 2005.
- [Booch 05] Booch, G., Rumbaugh, J., Jacobson, I.: *The Unified Modeling Language User Guide*, Second Edition, Addison-Wesley, 2005.
- [Brachman 04] Brachman, R. J., Levesque, H. J.: *Knowledge Representation and Reasoning*, Morgan Kaufmann (2004)
- [Heath 11] Heath, T., Bizer, C.: *Linked Data: Evolving the Web into a Global Data Space*, Morgan & Claypool Publishers (2011).
- [Parsons 90] Parsons, T.: *Events in the Semantics of English*, MIT Press (1990).
- [Pitler 08] Pitler, E., Raghupathy, M., Mehta, H., Nenkova, A., Lee, A., Joshi, A.: Easily identifiable discourse relations, In *Proceedings of COLING 2008* (2008).
- [Schank 77] Schank, R. C., Abelson, R.: *Scripts, Plans, Goals, and Understanding*, Lawrence Erlbaum Associates (1977).
- [Schema.org] Schema.org: Schema.org, <http://schema.org/>
- [Wang 14] Wang, Z., Zhang, J., Feng, J., Chen, F.: Knowledge graph embedding by translating on hyperplanes. In

Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (2014).

[山田 18] 山田隆弘: 世界知識の構造に基づいた談話理解モデル, 言語処理学会第 24 回年次大会, A2-3, 2018.