

パスウェイ知識表現における矛盾検証の理論的解析と技術開発

●福田 賢一郎¹⁾ ◆兼岩 憲²⁾

1) 産業技術総合研究所 2) 情報通信研究機構

<研究の目的と進め方>

疾病や生物プロセスなど、生命現象のメカニズムを対象とした知識処理では、論文やデータベースのコメント欄に自然言語で記述された知識の電子化手法が従来の課題であった。そのため、遺伝子機能を記述するためのオントロジーや、論文から該当知識を抽出するテキストマイニングなどの自然言語処理技術の研究開発が世界的に進められてきた。また、生命現象の分子メカニズムを記述するパスウェイデータベースについては、オントロジーで詳細に意味を定義したデータフォーマットを標準化させようという国際的な取り組みがすすんでいる。これらの成果により、近い将来、専門家によってキュレートされた知識、自然言語処理システムが自動抽出した知識、そして様々な実験技術から導出された知識、というように精度も性質も異なる情報が共通のプラットフォーム上で入手可能になることが期待される。このような状況を見据えて、今後は互いに整合性を持っていると保証されていない異なる出自のデータを対象とした知識処理技術の開発が必要となる。

特に複雑な情報構造を持つパスウェイデータベースについては、様々なレベルで矛盾・不整合を含むパスウェイの知識処理が望まれる。しかしながら、パスウェイ知識の矛盾検出については、知識表現の理論的な解析を含む基礎研究がほとんどなされていない。

本研究では複数オントロジー存在下での推論、パスウェイデータベースを用いた知識統合、新規知識の導出を目標として、パスウェイデータの矛盾検出に関する知識表現の基礎研究を展開する。

<研究開始時の研究計画>

研究代表者らが開発してきた「INOH」データベースおよびアノテーション用オントロジーをオントロジー記述言語に置き換える方法を検討するためにオントロジー記述言語としては以下の2つを想定した。

1) 研究分担者が開発しているソート論理や Description Logic のように概念階層を持つ論理を応用したオントロジー記述言語へ「INOH」データを変換する方法を設計する。

2) Web オントロジー記述言語 OWL への変換する方法を設計する。OWL での記述方法の設計は独自に行うのではなく、研究代表者が開発にたずさわっているパスウェイデータ・フォーマットの標準化をすすめている BioPAX の提供するフォーマットに基づいた開発を行う。

しかしながら、対象領域としているパスウェイデータには多数の概念とそのプロパティが関わり、またそれらの利用のされ方も複雑であるため、2006年度は生化学反応や生体高分子の修飾部位情報などを含む生化学反応レベルの詳細なパスウェイ情報の表現にはすでにオントロジー言語として確立されている OWL 言語での表現方法の検討をすすめることとした。論理による直接の表現は、生物プロセス間の関係の記述など、より抽象的なメカニ

ムの表現について2007年度に検討することとした。

<研究期間の成果>

本年度は主として以下の成果を得ることができた。1) イベント知識ベース構築のための形式オントロジーの開発。2) BioPAX オントロジーで記述されたデータをバックエンドデータベースに格納するための検索ライブラリ paxtools の開発。3) INOH データの BioPAX フォーマットでのリリース。

1) はイベント表現を蓄積する知識ベースの基盤となる上位オントロジーである。パスウェイは様々な要素が多様な関係性をもちながら細胞機能という現象を引き起こす分子メカニズムであるが、この現象の表現を一般化したものがイベントである。

```
Event
  NaturalEvent
    Occurrence1: {Time, Location}
    Occurrence2: {Object, Time, Location}
  ArtificialEvent
    Action1: {Agent, Object, Time, Location}
    Action2: {Agent, Time, Location}
    Action3: {AgentGroup, Time, Location}
  DynamicState
    ObjectChange: {Object, Time, Location}
    EnvironmentChange: {Time, Location}
  StaticState
    ObjectState: {Object, Time, Location}
    EnvironmentState: {Time, Location}
```

図1：イベントの分類

我々の提案するオントロジーはイベントの構成物と意味機能に着目することで、図1に示すようなイベントタイプを定義している。すなわち、イベントはそれを引き起こす主体や対象物の存在によって NaturalEvent と ArtificialEvent に分類される。NaturalEvent は行為者 (Agent) と対象物 (Object) を含まないことが特徴となっている。一方で、行為者によって実行される ArtificialEvent は、行為者と他の構成物との関係の違いで：行為者から行為者もしくは対象物に対する行為、自動詞的行為、複数の行為者による事象、の3つに分類される。提案しているイベントの上位オントロジーでは上記二種類のイベント以外にさらに DynamicState と StaticState が定義され、それぞれ時間と場所で変化する状態をイベントの低位概念としている。哲学などの関連研究では状態とイベントを別物とする考え方もあるが、解釈や視点によって状態もイベントと見なせることから提案手法では図1の構造をもたせている。DynamicState は対象や環境の状態が時間や場所で変化する振る舞いをしめし、StaticState は時間や場所の限定的な属性や性質を示している状態である。

また、意味機能による分類 (図2) とイベント間の関係要素を定義することで図3に示すように4つの因果関係を分類することができた。

EventSemanticFunctions

StateChange

TemporalExistenceChange

SpatialExistenceChange

CardinalityChange

Comparison

ObjectIdentificationChange

図2：イベントの意味機能タイプ

Causal Relations over Time

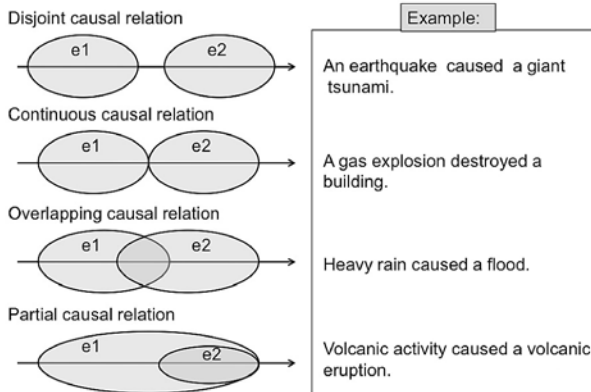


図3：4つの因果関係

2)についてはINOHのデータなど、BioPAX形式のパスウェイデータを扱うライブラリであるpaxtoolsで関係データベースをバックエンドに持たせるpersistence layerの開発を行った。

3)については、昨年度成果でもBioPAXオントロジーでは表現できないパスウェイ知識のOWLによる表現手法を開発したが、さらに転写制御メカニズムなどの表現方法をOWLにより表現し、BioPAX Level3の仕様に反映させた。INOH release 2.0はBioPAX Level3の正式アナウンスがまだなされていないため、従来通りLevel2でのリリース更新となっている。シグナル伝達パスウェイの表現などに対応するLevel3でのリリースは今年度末を予定している。

<国内外での成果の位置づけ>

イベントの表現は知識表現の大変深遠かつ基本的な問題であるが、本研究では構造表現に優れたソート階層を用いてイベントの構成物と意味機能の2つのオントロジー的な視点からイベントのタイプに着目し、分類を行った。Barry SmithのRelation Ontologyを修正することで因果関係のようなイベントの関係を定義している。

<達成できなかったこと、予想外の困難、その理由>

当初計画では、INOHデータを記述するためのオントロジー言語として、(Description Logicに基づく)Webオントロジー記述言語OWLとあわせて、ソート論理のように概念階層を持つ論理を応用した記述言語での記述を検討し推論メカニズムまでを開発する計画であった。しかしながら、対象領域としているパスウェイデータには多数の概念とそのプロパティが関わり、またそれらの利用のされ方も複雑であるため、論理による直接の表現は、整合性ルールや生物プロセス間の関係の記述など、より抽象的なメカニズムの表現について検討することとし、昨年度は生化学反応や生体高分子の修飾部位情報などを含む生化学反応レベルの詳細

なパスウェイ情報の表現について、すでにオントロジー言語として確立されているOWL言語での表現方法の検討をすすめた。今年度は抽象的なメカニズムの表現について研究を進め、パスウェイのメカニズム記述を抽象化した概念としてイベント知識を導入し、その上位オントロジーの研究にとりくむことでイベント知識ベース化に必要な論理的な基盤を構築した。

このようにパスウェイの扱う対象が広範であることから、一般化したイベント知識の論理的考察と具体的なデータの表現の研究を並行してすすめることとなり、論理から応用までをつなげる部分の研究にまで至ることができなかった。

<今後の課題、展望>

本年度の成果ではパスウェイデータベースを一般化した概念であるイベント知識ベースを構築するために必要となる上位オントロジーとその論理的形式化によるイベント知識の特徴付けを提案した。また、イベント知識による論理的推論の基盤を構築した。一方で、昨年度にはOWLオントロジーによる従来のパスウェイデータ表現の拡張を行った。今後の課題として、イベント間関係に対する推論メカニズムの設計とOWLなどで表現されたバイオのデータへの応用があげられる。

<研究期間の全成果公表リスト>

1) 論文/プロシーディング

- 0801291420
Kaneiwa,K.andIwazume,M.andFukuda,K.: An Upper Ontology for Event Classifications and Relations, Proc. of AI2007, Lecture Notes in Computer Science, 394-403 (2007).
- 0801291431
Fukuda,K.: INOH pathway database: Curation, Annotation, Integration, Proc of InterOntology08, in press.

2) データベース/ソフトウェア

- 0801291500
Paxtools Level2 persistence layer
http://www.inoh.org/inohblog/main/2007/10/paxtools_level2_persistence_la.html
- 0801291545
INOH release 2.0
http://www.inoh.org/inohblog/main/2007/11/inoh_release_20_has_been_relea.html