

意味情報ネットワーク構築に適した Context のモデル化

木下宏揚研究室 山下 祐平

(工学研究科 電気電子情報工学専攻 200770083)

1 はじめに

近年、検索技術の発達に伴いウェブ上に存在するような情報の相互利用が急速に進みつつある。現在、このような状況において情報を検索する方法として、一般に普及しているのはキーワードに基づいたテキストベースの方法である。しかしながら、キーワードに基づいたテキストの情報検索では、同じキーワードでもユーザが所望する意図がそれぞれ異なり、ユーザが満足するものを提供することができない。

これを解決するために意味情報に基づいた情報ネットワークの構築に関する研究が盛んに行われている。

意味情報に基づいた情報ネットワークは、意味を考慮した検索であるから、その精度は疑いの余地がない。しかしながら、正確な意味情報ネットワークを実現するための構築コスト、グラフ探索であるが故の探索のコストなどの難問もあり、一般には普及していない。

以上の問題を改善するためには、意味ネットワークの構築コストと探索コストの削減が不可欠となる。そこで、これらを実行するために Context という概念を定義し、適切な意味ネットワークへ分割し、分割した Context を弱い紐帯で結ぶといった手法が有効であると考えられる。この手法は、Context を分断してはいけないつながりまで分けてしまうと探索そのものができなくなるなど問題もある。しかしながら、所望する情報の意味同士が厳密につながっているとわかっているときには、Context に基づいた比較的狭い範囲を探索すればよいと考えられる。また、Context 間の意味関係を定義してあげると、弱い紐帯の役割を果たし、連想のような検索も行いう事ができると考えられる。

著者らは、21世紀 COE プログラム「人類文化研究のための非文字資料の体系化」[1]の一環として、民俗文化を知的文化遺産とし、研究成果を電子化、情報発信することを目指している。Dublin Core[2]を参考にした非文字資料に適したのメタデータの提案を最初に、意味情報を加味した非文字資料のデジタルアーカイブ化に関連する研究 [3][4][4]を行ってきた。その中で、Context の記述にはトピックマップが適しているとしてきた。[6]

そこで本稿では、トピックマップによる Context の記述を推し進め、意味情報ネットワークを構築する際に Context の適切な分割をどのように行えばいいのか、Context の特性を考察し実際にデータ構造を与えてみることでどのような利用法があり利点があるのかを示した。

また、弱い紐帯を利用した連想のような検索を可能とするには、Context 間の結びつきをどのようにモデル化するのかを考察し提案した。

以下、2節では、非文字資料と既存の意味に基づいた情報システムについて述べる。次に、3節では Context を導入したオントロジのモデルを提案する。4節ではモデルの利用法について述べる。最後に5節でまとめを行う。

2 Context の分析と提案モデル

2.1 Context の分析

Context は、直訳すると文脈といわれるように前後関係や背景、暗黙の前提条件などといった意味合いのものである。人工知能分野においては、意思伝達、言語学、形而上学などに属する部分と深い関係がある。

人間は、このような暗黙の了解としての Context を考慮することが得意である。この Context につい

A Modeling of Contexts suitable for Semantics Information network: Yuhei YAMASHTIA(Graduate School of Electrical, Electronics and Information Engineering, KANAGAWA University).

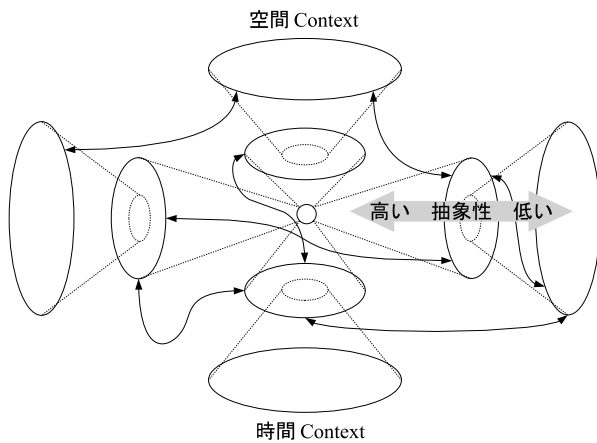


図 1: Context の軸と抽象度

て両者が合意していないとコミュニケーションもままならないことになる。

一方でコンピュータは、苦手である。例えば、検索を行う際に共通語彙が異なる概念として現れる場合も多いが、それは属する Context が違うためといえる。この場合、Context を考慮せずに処理を行うと見当違いの情報を返してしまう。

Context が分割されていれば、探索の際に行われる計算も少なくてすむ。それだけでなく、Context 間の関係についてメタデータを記述しておけば探索経路の優先度の決定に活用したり、Context は違うが意味構造がほぼ一致している意味グラフから推論を行い仮説を立てることが可能である。しかし、不適切な Context を与えると意味的に切り離してはいけな部分に分けてしまったり、本来探索が不要である Context を探索することが起こりうる。

このことから Context を適切に設定することは本方式においてきわめて重要な問題となる。ここでは、Context の主な種類、抽象度などについて考察する。

2.2 Context の関係性

Context は多く定義できる方がよいが、関係性が複雑になると計算量が増大する。そうすると Context に分割した当初の目的である探索の軽量化ができない。

2.2.1 Context の軸と抽象度

制約を加えないと Context は無数に定義できて混沌としてしまい、Context を用いていない状況と変わらなくなってしまう。そこで、図 1 に示すようにドメイン Ontology を多次元の軸に分離し、推移律が成立する Context 同士を従属関係にしたがって並べていくことでそれぞれの Context の抽象度を定義する。また、それらの組み合わせで任意の Context を表現することにより、それぞれの軸での特徴から Context の関係を導くことが可能となる。

2.2.2 Context の組み合わせ

図 1 の各軸に属する Context においてそれらを組み合わせることができれば、任意の Context を作るのと同様になる。ここでは、Context の組み合わせ方について簡単に述べる。

まず、組み合わせの方法は既存のモデルと同様の論理和、論理積などを使った組み合わせに限定する。これは、集合の側面をもつ Context の特性を生かしたものである。ここでなぜ意味関係を使用しないかということ Context 自体がメタ的な概念であるから意味関係を考慮するよりも処理量の削減を重視すべきという理由からである。

Context を任意に定義した場合と同じようにさまざまな Context を定義可能だが、Context が増えていくことによる混乱は、Context の軸の数が Context の数に比較してかなり少ないことによって抑えられる。

3 提案システム

Context は多く定義できる方がよいが、関係性が複雑になると計算量が増大する。そうすると Context に分割した当初の目的である探索の軽量化ができない。

このことをふまえ、以下のモデルを提案した。

- (1) 民俗資料などの各ドメイン Ontology を構築する。
- (2) Ontology で定義した概念を分析し、概念の依存関係を元に独立した概念を決定する。
- (3) (2) で定義された独立した概念を属性として保持し使用頻度の高い Context を規定する。

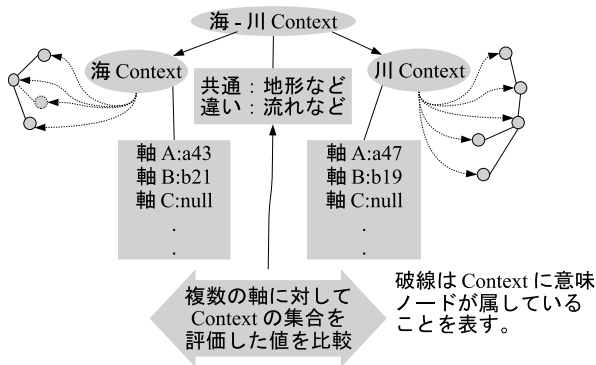


図 2: 提案する Context 構築のためのデータ構造

(4) (3) で定義された Context に現れる特筆すべき関係を Context 間の差延として表現する。

これは、言い直すと Context に抽象度と組み合わせという論理を導入することである。

3.1 Context のデータ構造

図 2 は、今まで述べてきた Context を実際にデータ構造として記述した場合である。

まず、海に関係している Context と、川に関係している Context をそれぞれ Context のノードとして設定する。それらの Context が、図 1 の階層化された軸のどの Context を組み合わせて作成されているのかを調べる。組み合わせ元の Context を Context ノードの内部出現のリストとして保持する。これにより Context 自体は表現されたことになる。

次に、作成した Context の特性リストからある軸について一致している場合や、逆の特性を指している場合など特筆すべき関係を持っている場合のみ、Context ノードを接続し、各軸における重要な Context 関係を内部出現として保持する。

このように Context 間の関係を示すノードに記述しておけば、探索や推論などでうまくいかなかった Context を置き換えるときに Context 間の関係を表すノードを参照するだけでよくなる。Context の関係が既知の場合には、未知の場合に対して計算量は大きく削減されて探索の精度も向上するというメリットがあり、Context のノードを作成して軸に属している Context の組み合わせを保持することは割に合うということがいえる。

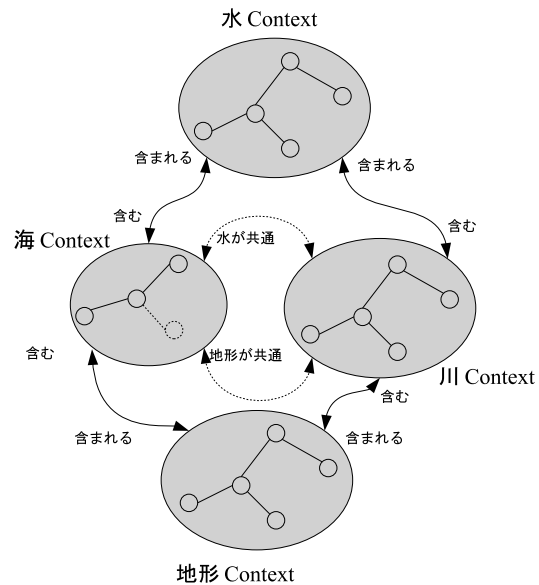


図 3: Context 間の類似度

3.1.1 Context の類似度

軸に分離された Context の組み合わせによって作られた Context 同士の関係は、Context の軸の概念距離から導出することが可能である。特に、類似した性質を持っている Context については、連想検索などに利用することが可能である。

さらに、未知の知識に関する仮説を類似する Context から推論し提供することができる。図 3 は、Context 同士の関係が存在するときにそれがどのような構造として現れるかを示したものである。図では、海 Context と川 Context は水を含んでいるものでも地形であることを示している。それらの関係は Context の軸と抽象度で説明したように推移可能な Context 同士の概念距離やその共通祖先によって定められる。ここでは、海 Context と川 Context がともに共通の祖先である水 Context および地形 Context をもつことから海 Context と川 Context には共通属性があるという Context 間の関係を示している。

4 提案モデルの利用例

ここまでに示してきた Context の分割を活用した探索時の優先探索経路の決定アルゴリズムを図 4 に示す。

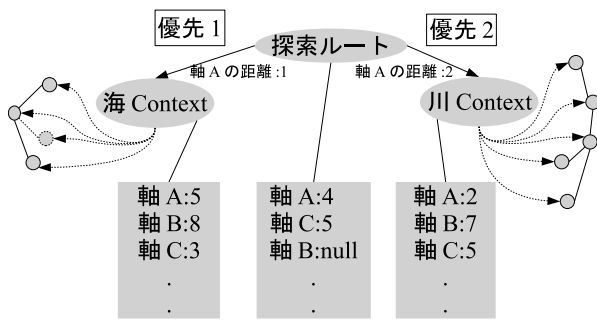


図 4: Context の検索経路の優先度

図 4 は、意味グラフのパターンマッチングを行う際に着目しているカレントノードの優先度の変化を示している。

まず、ユーザが生成したクエリの意味グラフからその Context の特性を各軸についてたどり、Context の特性リストとする。これを Context ノードが保持している Context の属性と比較する。その際、優先度の高い軸（図では軸 A）から順番に比較して概念距離に近い Context を重要な検索対象であると考えられる。したがって、クエリから、は少なくとも軸と軸の優先順位を導かなければならない。しかし、それは、意味ネットワークグラフと同じものを使うことができる野でそれほど問題にはならない。

意味に基づいた情報検索では、次に探索するノードの順番が非常に重要になる。例えば、探索中のノードの隣接ノードに解があるにもかかわらず膨大な隣接ノードをもつ別のノードを先に調べてしまい、それらのノードを探索し終えてから探索するのでは探索コストが非常に異なる。このことから、曖昧な概念集合としての Context の探索の優先順序を簡単に導出できる提案方式のデータ構造は、探索の応答性をあげるのに優れているといえる。

5 まとめ

本稿では、意味に基づいた検索を行う際に有益であると思われる Context について考察してきた。これまでは主として意味グラフのパターンマッチングにおける探索の計算量を削減することを主張してきたのに対して、本稿では Context の特性を考察し、既存の情報システムにはなく意味情報に基づいた情報

ネットワークに独自のグラフ探索や推論などに活用できる適切な Context 分割を行う方法について具体例を交えて示した。

まず、Context の適切な分割を目的とした Context の概念の軸と抽象度の導入である。Context の分類を推移律が成り立つ軸にたいして抽象度を変化させることができる Context の組み合わせで記述することで、Context 間の関係性がより明確になり、より効率的な検索や推論を行うことができる。Context に軸と抽象度という仕組みを取り入れることで、さまざまな問題が解決されると考えられるが、Context の概念階層の設計には、Ontology の構築と同じように厳密な意味論理を損なわないように設計しなくてはならない。今後の課題は、これらの Context を適切に設定したり、自動で調整を行えるような仕組みを構築することである。

参考文献

- [1] 神奈川大学 “21 世紀 COE プログラム人類文化研究のための非文字資料の体系化” <http://www.himoji.jp/>
- [2] Dublin Core Metadata Initiative., <http://www.dublincore.org/>
- [3] 木下慶子, 稲積泰宏, 木下宏揚, 森住哲也 “COE における非文字資料の共有と流通”
- [4] 木下慶子, 稲積泰宏, 木下宏揚, 森住哲也 “非文字資料に適した Ontology 構築”
- [5] 木下慶子, 村上敦志, 稲積泰宏, 木下宏揚, 森住哲也 “デジタルアーカイブにおける Ontology の活用”
- [6] 山下祐平, 木下宏揚, 森住哲也 “トピックマップを用いた非文字資料における Context の表現”