

## SAT ソルバを利用したオントロジー上の概念間距離の計算

長野 伸一

稲葉 真純

服部 正典

(株) 東芝 研究開発センター

## 1 はじめに

オントロジーを利用した情報抽出や文書処理に関する研究が進められている。例えば、溝口らは、文書に出現する語彙をオントロジーの概念と対応付け、概念間の距離をもとに文書間の類似度を評価する手法 [1] を提案している。一方、オントロジー応用において、オントロジー品質の定量化が課題となっている [2]。Tartirらは、オントロジーの構造的特性に着目した品質メトリックスを提案している [3]。これらの手法では、オントロジー上の概念間距離の計算が、基本計算の 1 つとして利用されている。オントロジー上の概念間距離の計算とは、オントロジーの概念、関係それぞれをノード、エッジとするグラフとみなし、グラフ上のノード間の距離を求めることをいう。

本稿では、OWL で記述されたオントロジーを対象とし、指定した 2 つの概念間の距離を求める問題について論じる。Weinstein らは、メモリ上にグラフを構築し、グラフ上のパス探索を行うアルゴリズムを提案している [4]。しかし、概念間距離の計算はオントロジーの規模に対して指数時間を要する。本稿では、オントロジーの階層構造を無向グラフとみなし、その構造をブール制約式として変換する手法を提案する。この変換により、概念距離計算をブール式の充足可能性判定問題とみなし、SAT ソルバを利用して求めることができる。メモリ上にグラフを構築する必要がないため、概念間距離を効率良く計算することが可能となる。

## 2 SAT 問題

SAT 問題とは、制約式が与えられたときに、それに含まれるブール変数に対して真または偽をうまく割り当てて、制約式を真にできるかどうかを判定する問題である。SAT 問題も、理論上はブール変数の数に対して指数時間の計算量を必要とする問題であるが、近年 SAT 問題を高速に解くアルゴリズムとツール (SAT ソルバと呼ぶ) の研究開発が進んでおり、様々な分野へ応用した成功事例が数多く報告されている [5]。

本研究では、Pseudo-Boolean (PB) と呼ばれる形式の論理式を用いる。PB 制約式は、ブール変数  $x_i \in \{0, 1\}$

に自然数の係数  $C_i$  を付与した線形不等式で、次の形式で表される。

$$C_0 \cdot x_0 + C_1 \cdot x_1 + \dots + C_{n-1} \cdot x_{n-1} \geq C_n$$

変数  $x_i$  がある割り当てのもとで真となる場合、対応する係数  $C_i$  は活性化されるという。左節の活性化された係数の和が、右節の係数  $C_n$  より大きい、または一致する場合、その PB 制約式は充足されるという。

PB 制約式の充足可能性判定は、PB-SAT 問題と呼ばれる。PB-SAT 問題は、SAT 問題を拡張して最適化問題を扱うようにしたものである。すなわち、目的関数がブール変数の線形の式で与えられたとき、全ての制約式を真にするような、ブール変数への割り当てのうち、目的関数の値を最大化あるいは最小化するような割り当てを求める問題である。SAT 問題が、制約式の充足可能性を求めるのに対し、PB 問題は目的関数の元でブール変数への割り当てを決定する問題である。代表的な PB-SAT ソルバには、SAT4J<sup>\*</sup>、MiniSat+<sup>†</sup> などがある。

## 3 OWL から PB 制約式への変換

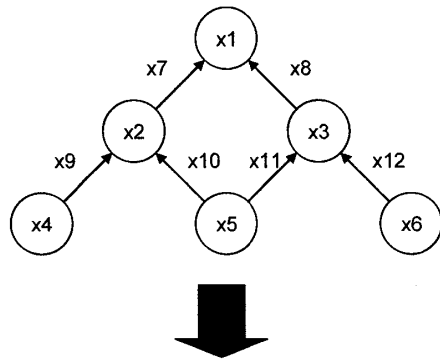
提案法は、OWL 形式のオントロジーデータ、および 2 つの概念を入力とし、与えた概念間の距離を計算する。オントロジーの概念間の関係を表す PB 制約式、および与えた概念間の最短経路を求める目的関数を生成し、PB-SAT ソルバを利用して、目的関数の値を最小化するような PB 制約式のブール変数への割り当てを求める。もし PB 制約式が充足可能であれば、概念間のパス長を距離として出力する。充足可能でなければ、経路が存在しない。提案法の特徴は、オントロジー上の最短経路を求めるグラフ探索問題を、PB 制約式の最適化問題として解く点にある。これにより、メモリ上にグラフを構築することなく、概念距離を計算することができる。

図 1 に変換の例を示す。オントロジーは、6 つのクラス、および 6 つの関係から構成されている。提案法は、各クラスを PB 制約式の集合へと次のように符号化する。まず、クラス、関係それぞれに対して、ブール変数  $x_i$  をラベル付けする。各クラスに対する PB 制約式

An Approach to Conceptual Distance Measurement on Ontology  
Shinichi NAGANO, Masumi INABA, Masanori Hattori  
Corporate R&D Center, Toshiba Corporation

<sup>\*</sup><http://www.sat4j.org/>

<sup>†</sup><http://minisat.se/>



```

* #variable= 12 #constraint= 8
* objective function
min: +1 x7 +1 x8 +1 x9 +1 x10 +1 x11 +1 x12 ;
* constraints
+2 ~x1 +1 x7 +1 x8 = 2;
+2 ~x2 +1 x7 +1 x9 +1 x10 = 2;
+2 ~x3 +1 x8 +1 x11 +1 x12 = 2;
+1 ~x4 +1 x9 = 1;
+1 ~x6 +1 x12 = 1;
+1 ~x5 +1 x10 +1 x11 = 1;
* source node
+1 x4 = 1;
* destination node
+1 x6 = 1;
    
```

図 1: OWL から PB 制約式への変換の例

は、そのクラスが最短経路上に存在するか否かに基づいて、次の2つの場合の論理和として表される。(a) 着目しているクラスが、あるクラス対を結ぶ経路上に存在する。(b) 着目しているクラスは、どのクラス対を結ぶ経路上にも存在しない。ここで、例として、 $x_2$  とラベル付けされたクラスに着目しよう。クラス  $x_2$  に関する符号化は、図2に示したように、次の4つの場合に分類される。(a1)  $x_2$  は、2つのエッジ  $x_7$  と  $x_9$  を通る経路上に存在する。(a2)  $x_2$  は、2つのエッジ  $x_7$  と  $x_{10}$  を通る経路上に存在する。(a3)  $x_2$  は、2つのエッジ  $x_9$  と  $x_{10}$  を通る経路上に存在する。(b1)  $x_2$  は、どのクラス対の経路上にも存在しない。この結果として、 $x_2$  に対する PB 制約式は、上記4つの場合の論理和として、次のブール式で表される。

$$(1 \cdot x_7) + (1 \cdot x_9) + (1 \cdot x_{10}) + (2 \cdot x_2) = 2$$

ここで、ブール変数  $x_i$  の値が1の場合、変数  $x_i$  が表すノード、またはエッジが最短経路上に存在することを意味する。

次に、目的関数は、オントロジー上の任意のクラス対の最短経路を見つける PB 制約式として、次のように表される:

$$\min(w_0 \cdot e_0 + w_1 \cdot e_1 + \dots + w_{n-1} \cdot e_{n-1})$$

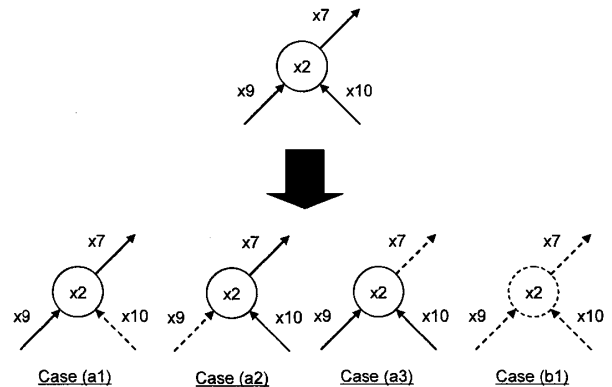


図 2: クラスから PB 制約式への符号化ルール

ここで、 $e_i$  はエッジを表し、 $w_i$  は自然数の係数で、エッジ  $e_i$  の重みを表す。図1のオントロジーの場合、エッジの重みは全て1と仮定すると、次の目的関数が生成される。

$$\min(1 \cdot x_7 + 1 \cdot x_8 + 1 \cdot x_9 + 1 \cdot x_{10} + 1 \cdot x_{11} + 1 \cdot x_{12})$$

変換の結果として、1つの目的関数、クラスに対して計6つの PB 制約式、および指定した2つのクラス ( $x_4$  と  $x_6$  が与えられらと仮定する) それぞれに対する PB 制約式が生成される(図1を参照)。PB-SAT ソルバにより、これらの PB 制約式は充足可能であると判定され、 $x_4$  と  $x_6$  の最短経路  $x_9 x_7 x_8 x_{12}$ 、および距離4が出力される。

#### 4 まとめ

本稿では、オントロジー上の概念間距離を、PB-SAT ソルバを利用してシンボリックに計算する手法を提案した。これにより、メモリ上にオントロジーのグラフを明示的に構築する必要がなく、効率良く計算することが可能となる。今後は、提案法の定量的評価を行う実験を実施し、有用性を検証する。

#### 参考文献

- [1] Y. Mizoguchi-Shimogori, et al., TV navigation agent for measuring semantic similarity between documents, Proc. of AWeSOMe, 2007.
- [2] A. Gangemi, et al., Modelling ontology evaluation and validation. Proc. of ESWC, 2006.
- [3] S. Tartir, et al., OntoQA: Metric-based ontology quality analysis, Proc. of IEEE ICDM KADASH Workshop, 2005.
- [4] P. C. Weinstein, et al., Comparing concepts in differentiated ontologies, Proc. of KAW, 1999.
- [5] SAT Conference, <http://www.satisfiability.org/>