

# 頻出議論部分木マイニング

## Frequent argument subtree mining

糸谷拓人\*<sup>1</sup>

Takuto Itoya

澤村一\*<sup>2</sup>

Hajime Sawamura

\*<sup>1</sup>新潟大学大学院自然科学研究科電気情報工学専攻  
Graduate School of Science and Technology Niigata University\*<sup>2</sup>新潟大学自然科学系  
Institute of Science and Technology Niigata University

In this paper, we propose a method for mining frequent structures from the data consisting of a set of argumentation trees. Concretely speaking, a set of trees whose parents are conclusions and children are premises of each argumentations is constructed and the frequent subtree mining algorithm called SLEUTH is applied to it. At this time, mining of argumentation is to be possible by putting a same label to nodes that show a same content.

### 1. はじめに

議論は、日常生活はもちろん政治、経済、司法、あらゆる学問分野で行われている。議論で主張する際、その主張の根拠となる情報は、文献や他者が述べたことを参考にする事が多い。一般的により多くの文献、人によって述べられている情報は、説得力を持つ。これは議論においても同様で、膨大な量の議論情報に頻出する情報を探すことができれば、議論において主張の根拠を強めることができるのではないかと考えた。しかしながら、このような議論情報を我々あるいは知識を利用するソフトウェアシステムにおいて有効利用しようとする研究は、ほとんど行われていない。澤村研究室では、議論情報のマイニングをする試みが行われてきたが、自然言語により表現された議論を対象としたマイニングの研究は行われていなかった。しかし、情報の大半は自然言語で表現されているため、有用な情報を得るには自然言語により表現された議論の集合を対象としたマイニングが必要である。

本論文では、自然言語で表現された議論のデータベースから有意な議論情報をマイニングする方法を提案する。議論は木構造で表現できるので、部分木の中でも特に誘導部分木や埋め込み部分木をマイニングできる頻出部分木マイニングアルゴリズムとして SLEUTH[1] を採用し、意味ある有用な部分木をマイニングする方法を与える。

### 2. 頻出議論部分木マイニングの手法

#### 2.1 マイニングのプロセス

1. 各議論を議論木の形式にデータクレンジング
2. 議論木を Araucaria[2] で描写し,AML ファイルを作成
3. 「頻出部分木マイニングアルゴリズム」を利用した頻出構造の出力
4. ユーザーによる出力された結果の検証

#### 2.2 議論木

本論文で扱う議論木は、議論における論証をデータクレンジングのため木に変換したものである。論証とは、前提群から結

論が真であることが導き出せることを主張したものである。議論木の親子関係は、親が結論、子が前提という構造である。

結論、前提の各命題にラベル付けを行い、同じ内容を示す命題には同じラベルを付けるようにする。これにより、議論のマイニングが可能になる。議論を木に変換する際、各ノードにはそのラベルのみが入る。

#### 2.3 Araucaria -議論分析ツール-

Araucaria は、議論を分析するソフトウェアツールである。簡潔なインターフェースを使って、議論のダイアグラムを作成できる。議論木の記述に Araucaria を使う理由は、木構造を入力することが可能であり、また AML というファイル形式で出力できるからである。議論は木構造で表現できるので、簡単に木構造を入力できるツールを使用する。また、頻出部分木アルゴリズムを利用するためには、木構造を文字列に変換する必要がある。そのため、XML 形式のデータである AML はフォーマットが決まっているので、一括したデータの変換を行うことが容易である。

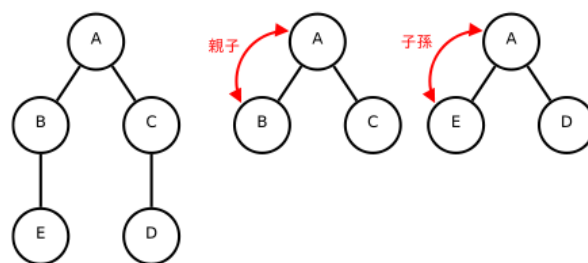
#### 2.4 部分木

データマイニングで出力される頻出な部分構造は 2 種類存在する。2 種類の部分構造について簡単に記述する (図 1)。

誘導部分木: 親子構造を保持している部分構造

埋め込み部分木: 子孫構造を保持している部分構造

部分構造の包含関係は、誘導部分木  $\subseteq$  埋め込み部分木となる。



1. もとの木

2. 誘導部分木

3. 埋め込み部分木

図 1: 部分構造の例

連絡先: 糸谷拓人, 新潟大学大学院自然科学研究科電気情報工学専攻, 新潟県新潟市西区五十嵐2 の町8 0 5 0, itoya@cs.ie.niigata-u.ac.jp

## 2.5 頻出部分木マイニングアルゴリズム

文字列に変換した論証の集合からよく現れる構造を探すために頻出部分木マイニングアルゴリズムを利用する。本論文では、「SLEUTH」というアルゴリズムを利用する。このアルゴリズムの主な特徴は、入力兄弟関係に順序が定義されていない木(無順序木)の集合であり、出力は「埋め込み部分木」である。また、オプションで「誘導部分木」も出力することが可能である。

## 3. データマイニングの応用例

議論のマイニングする際、議題を統一した方が出力結果に特徴が出やすい。そこで本論文では、まず神の存在証明という議題で、次に日本の神観という議題でマイニングを行う。

### 3.1 神の存在証明

神の存在証明とは、主に中世哲学における、理性によって神の存在する根拠を提示することである。神の存在証明は、古今東西で数多く議論されてきたため、注目し、議題として選択した。データベースは、インターネット上の様々なサイトに記述されている証明を参照した [1]。

今回用意したデータベースの規模は、議論木が 14 個、ノードが 98 個でラベルが 66 種類である。最小サポートを 2 個に設定 (2 個以上の部分木があれば頻出とする)。まず、埋め込み部分木のマイニングを行う。頻出部分木は、全部で 23 個出力された。その中でノードが 3 個以上出力された構造を図 2 に示す。図の実線は親子関係、点線は子孫関係をそれぞれ表している。

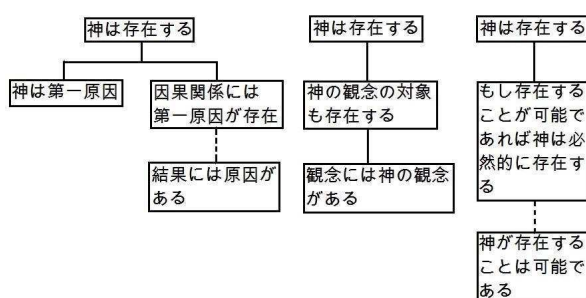


図 2: 埋め込み部分木

次に、誘導部分木をマイニングする。実験条件は埋め込み部分木の時と同じとする。頻出部分木は、全部で 18 個出力された。その中でノードが 3 個以上出力された構造を図 3 に示す。

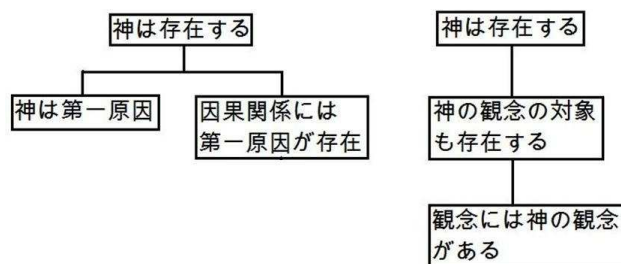


図 3: 誘導部分木

埋め込み部分木と誘導部分木を比較すると、抽出された頻出部分木の総数や内容にはあまり大きな違いは見られなかった。これは議論木の高さが低いものがほとんどであったためと考えられる。出力された部分木を見ると、神は完全 (最高・最大・第一原因) ということを前提とし、それをういて証明している論証が大半であった。つまり神は全能者または万物の創造主であるということとは有用な知識であると言える。

### 3.2 日本の神観との比較

3.1 節では、出力結果から神の存在証明のデータベースは、西洋的な一神教の知識が多かった。そこで日本固有の多神教の宗教である神道についても検証を行い、結果を比較する。

神道において、神は主に以下のような側面がある。

- 自然物や自然現象を神格化した神
- 思考・災いといった抽象的なものを神格化した観念神
- 古代の指導者・有力者などを神格化した神

これらについて論じている議論をデータベースとして、マイニングを行う。今回用意したデータベースの規模は、議論木が 6 個、ノードが 23 個でラベルが 17 種類である。最小サポートを 2 個に設定。埋め込み部分木でノードが 3 個以上の構造の出力結果を図 4 に示す。

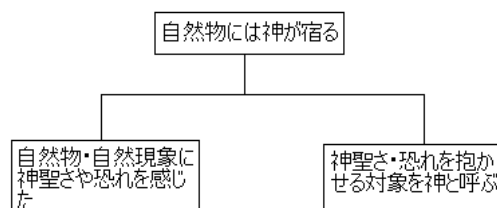


図 4: 神道のマイニング結果

出力結果から神道において神は畏怖の対象で、人々は人智を越えたものを神と呼び、崇めていたということが読み取れる。神の存在証明の結果と比較すると、神の存在証明では神は万物の創造者とされているので神は唯一無二の存在であることが分かる。それに対し、神道では神は畏怖の対象であるので、その対象は雷や嵐などの自然現象から死者の霊的なものまで無数に存在する。これらから、それぞれの信仰の対象がわかり、何を神と呼んでいるのかが分かる。

## 4. まとめ

本論文では、自然言語で表現された議論のデータベースから有意義な議論情報をマイニングする方法を提案した。検証の結果、議論に頻出するパターンを抽出し、知識として得ることは十分可能である。より多くの議題でマイニングを行えば、幅広い知識を得ることが可能となる。

## 参考文献

- [1] 頻出議論部分木マイニング, 糸谷拓人, [http://www.cs.ie.niigata-u.ac.jp/Paper/Storage/graduation\\_thesis\\_itoya.pdf](http://www.cs.ie.niigata-u.ac.jp/Paper/Storage/graduation_thesis_itoya.pdf)(2011).
- [2] Araucaria version3.1-USER MANUAL-(2008).