

## Web上の議論に対する頻出部分木マイニング

## Frequent subtree mining for arguments on the Web

東條亮平\*<sup>1</sup>      澤村一\*<sup>2</sup>  
Ryouhei Toujou      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 recent years, SNS, Blog and Twitter are popular in the world. The information communicated by people around the world in place on the Web, have a argumentation format which consists of its conclusion and rationale basically. The frequent argumentation in place on the Web is based on the thought that people have in common. In the previous study, Itoya proposed a method for mining frequent structures from the data consisting of a set of argumentation trees, and he did mining for proof was done by philosophers such as Thomas Aquinas. But mining for argumentation on the Web has not been done yet.

In this paper, we do mining for argumentation on the Web, and we consider results of mining.

## 1. はじめに

近年、蓄積された大量のデータから有用な情報を取り出す技術としてデータマイニングが注目されている。データマイニングの研究は現在ビジネス分野や科学技術分野などのさまざまな分野で行われている。議論の研究にもマイニングが利用出来ると考え、我々は議論をマイニングする試みを行っている。糸谷 [1] は自然言語の議論を木構造で表現して頻出部分木マイニングを行う手法を提案し、トマス・アキナスなどの哲学者により行われた証明を対象にマイニングを行った。しかし、Web上で行われている議論を対象としたマイニングは行われていない。Web上で行われている議論を対象にすることにより、現代の人が重要視している議論が発見出来ると考えた。

本研究では、"Justice has been done."とアメリカのオバマ大統領が発言したことに対して、Web上で行われている議論を日本、アメリカの二つの文化圏ごとにデータベースとして用意し、それぞれに対して議論の共通する構造をマイニングする。そして、マイニングの結果を考察する。

## 2. 議論のマイニング

### 2.1 Web上の議論

まず、本研究における議論とは論証のことである。Web上ではSNS、ブログ、ニュースのコメント欄などを利用し、世界中の人々が自分の主張を述べている。

### 2.2 議論木

議論は木構造で表すことが出来る。ここでいう議論とは論証のことである。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。

### 2.3 Araucaria-議論分析ツール

Araucaria は議論を分析し、ダイアグラムを作成するソフトウェアツールである。Araucariaを使うことにより自然言語で記述された議論から議論木を作成することが出来る。また、分析した議論をXMLに基づいたAML(Argument Markup Language)というフォーマットで保存することが出来る。

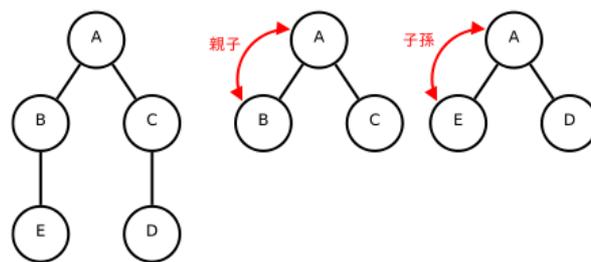
### 2.4 部分木

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  $\supseteq$  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

### 2.5 頻出部分木マイニング

議論は木構造で表現できるので頻出部分木マイニングを利用することにより頻出する議論の構造を探す。頻出部分木マイニングアルゴリズムはいくつか提案されている。入力が無順序木であることと出力が誘導部分木であることを理由に本研究では、SLEUTHアルゴリズムを利用する。

連絡先: 東條亮平, 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 8050, r-toujou@cs.ie.niigata-u.ac.jp

### 3. Web上の議論に対するマイニングの実験

#### 3.1 目的

Web上の議論をマイニングすることにより現代の人の思想や信念に基づく議論を発見できることを確かめる。

#### 3.2 方法

マイニングする議論の範囲を決めなければ頻出する部分議論は見つかりにくいと考えられるので実験に用いる議論のテーマを定める。実験にはオバマ大統領の発言”Justice has been done.”について行われている議論を用いる。これは、オサマ・ビンラディンがアメリカ軍により殺された後にオバマ大統領が会見で発表したものである。この発言をニュースで聞いて”Justice”と発言したことに疑問を感じ、議論のテーマとして選択した。

日本とアメリカの異なる文化圏の議論をそれぞれマイニングし、結果を比較し、考察する。実験に用いる議論のデータは、Web上のニュースのコメント欄やBlogから参照した。

#### 3.3 実験手順

1. Web上の議論から Araucaria を使い議論木を作成
2. 議論木のノードにラベルをつける  
この時同じ意味を表す命題には同じラベルをつけることで、議論のマイニングが可能になる。
3. マイニングアルゴリズムを利用して、頻出である部分木を出力
4. 頻出部分木から共通の思想、信念を考察する

#### 3.4 データベース

データベースの規模は、日本人が行っている議論の議論木が27個、ノードが98個でラベルが63個で、アメリカ人が行っている議論の議論木が24個、ノードが80個でラベルが38個である。

#### 3.5 実験結果

実験では、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)。

##### 3.5.1 日本人の議論

日本人の議論では頻出部分木は、全部で30個出力された。その中でノードが2個以上の頻出部分木を図2に示す。また、頻出部分木の下に頻出した数を示す。

ここで、点線は埋め込み部分木を表している。

##### 3.5.2 アメリカ人の議論

アメリカ人の議論では頻出部分木は、全部で22個出力された。また、頻出部分木の下に頻出した数を示す。その中でノードが2個以上の頻出部分木を図3に示す。

#### 3.6 考察

日本人の議論では、オバマ大統領が”Justice”と発言したことに対して否定的な議論が頻出した。その中でも「正義はなされていない」という結論を導く議論が特に頻出した。それらの根拠には「裁判をしなければ正義でない」「罪を明確に証明していない」などがあり、日本人は司法を重要視する思想を持っていると考えられる。アメリカ人の議論では、肯定的な議論と否定的な議論の両方が頻出した。肯定的な議論では根拠に「ビンラディンは何千人も殺した」とあり、これは旧約聖書や新約聖書の各福音書にも同様の記述がある「目には目で、歯には歯で」というハンムラビ法典のタリオの法に由来する思想に基づいていると考える。否定的な議論では日本人の議論と同じ頻出

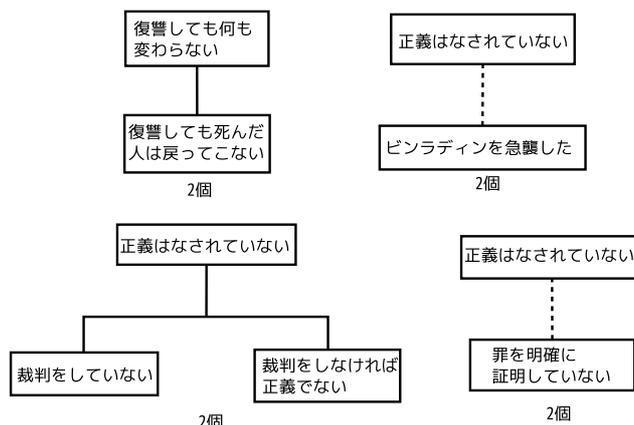


図 2: 日本人の議論での頻出部分木

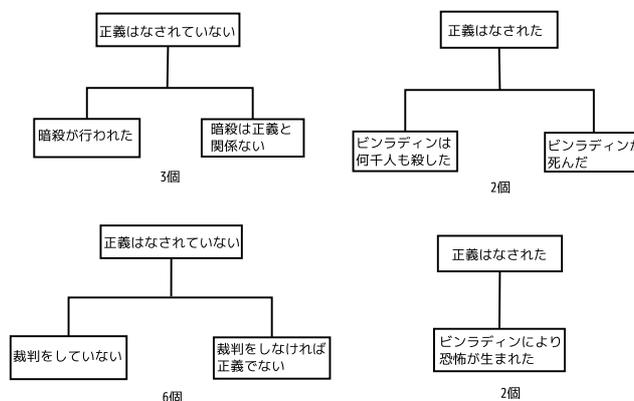


図 3: アメリカ人の議論での頻出部分木

部分木が頻出した。つまり、アメリカ人も日本人と同様に司法を重要視する思想を持っていると考えられる。これは日本人とアメリカ人に共通の思想であると言える。

### 4. まとめ

本研究では、Web上で行われている異なる文化圏の議論に対し、それぞれ頻出部分木マイニングを行った。

実験の結果、Web上で行われている議論をマイニングすることにより、現代の人の思想や信念に基づく議論を発見することが十分可能である。また、異なる文化ごとに議論をマイニングすることにより、それぞれの文化の特徴を表す議論を発見できる。このことより、Web上で行われている議論をマイニングすることは人文系の研究にも寄与する可能性があると言える。

### 参考文献

- [1] 頻出部分木マイニング, 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-(2006).