

令和 5 年 5 月 4 日現在

機関番号：12601

研究種目：基盤研究(C)（一般）

研究期間：2020～2022

課題番号：20K11931

研究課題名（和文）元テキストの論理構造を適切に反映した議論グラフの作成手法の構築

研究課題名（英文）Argumentation Graph Construction Method Reflecting the Logical Structure of Natural Language Arguments

研究代表者

山口 和紀（Kazunori, Yamaguchi）

東京大学・大学院総合文化研究科・名誉教授

研究者番号：80158097

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：近年、自然言語で書かれた議論を議論グラフで表わし、その議論グラフを分析することが盛んに行われている。しかし、自然言語の議論の構造を反映した議論グラフを作ることは難しい作業である。本研究では、自然言語の意味の一つである類似度を取り込んだ議論グラフの定式化、およびその評価手法と分析手法を構築した。理論的な分析から評価手法が良い性質を持つことを示した。実際の議論グラフに適用した結果から、評価手法と分析手法が有用であることを示した。構築した評価手法は拘束条件をコストとして表すことに基づいており、新たな意味の追加が容易であり、今後の発展の余地が大きい。

研究成果の学術的意義や社会的意義

議論は、科学技術で結論を導いたり、政策決定で意思決定のために使われるなど人間の思考における重要な道具であり、議論を適切に進めるための支援は重要である。本研究は、自然言語で書かれた議論を議論グラフとして表現し、その表現の上で評価・分析する手法を与えるものであり、そのような支援の重要な基礎となる。攻撃や支持の構造に関しては論理的な性質に基づいた予測や分析が可能であり、自然言語で表現された言明の意味の一つである類似度に関しては、攻撃・支持と組み合わせた評価が可能である。これにより様々な観点から議論グラフの表現が適切であるかの判断を支援することを可能とした。

研究成果の概要（英文）：In recent years, it has been popular to represent arguments discussed in natural language by argumentation graphs and to analyze them. In this research, we formulated an argumentation graph incorporating similarity, which is one of the meanings of the natural language, and constructed its evaluation and analysis methods. From the result of applying them to the actual argumentation graph, it was shown that the evaluation method and the analysis method are useful. The constructed evaluation method is based on expressing constraints as costs, making adding new meanings easy and providing room for future extension.

研究分野：情報科学

キーワード：議論グラフ

1. 研究開始当初の背景

議論は、科学技術で結論を導いたり、政策決定で意思決定のために使われるなど、人間の思考において重要な道具である。近年、従来自然言語で書かれた議論を議論グラフで表現し、その議論グラフを分析することが盛んに行われるようになってきた。この方法では、自然言語で書かれた議論のレトリックなどを排除し、議論中の最小要素(言明)間の攻撃や支持などの関係(主張)に着目して議論の客観的な分析が行えるという利点がある。一方で、自然言語で書かれた議論を議論グラフとして形式化するとき、議論中の誤った部分を言明と捉えて議論グラフを作ってしまうことで、議論の意味からはずれた分析をしてしまう危険性も指摘されている。議論グラフの専門家が意味を考えて議論を分析すれば、議論を正しい議論グラフで表わすことができるが、議論グラフを用いた分析を非専門家が利用できるようにするためには、議論の構造を正しく反映した議論グラフを作ることへの支援が必要とされている。

2. 研究の目的

自然言語で書かれた議論の構造を正しく表した議論グラフを一般に作成可能とするために、本研究では、議論の構造を反映しやすい議論グラフの構造を与えることと、議論グラフの言明と主張(この2つを議論要素と呼ぶ)がその議論グラフでどの程度成立しているかを求める評価手法および主張がどの程度満たされているかなどの分析指標に基づいた分析手法を考案することを目的とした。議論グラフの構造はこれまで検討されている議論グラフの多くを表現できるとともに、自然言語の意味の一部を取り入れられるようにする。評価手法は、このように定義した議論グラフの議論要素の値を計算でき、将来の拡張にも耐えられる一般性の高いものを構築することを目的とした。分析手法としては、議論がどの程度収束しているかや他の部分と整合性の低い主張の発見に繋がるものを提案することを目的とした。

3. 研究の方法

本研究では、議論グラフを定式化し、議論グラフの評価手法と分析手法を考案した。そのために過去の研究を参考にしながら、良い性質を持つものが得られるまで試行錯誤を繰り返した。自然言語で書かれた議論から議論グラフが正しく生成できるようにするためには、まず、十分な表現力のある議論グラフの構造が必要となる。過去の研究から、攻撃と支持、再帰(攻撃や支持の先がまた主張であるような主張)、推定的証拠の原則(攻撃されていない議論要素の値は1とする)を表現する必要があることがわかった。これらでは表現しきれない自然言語で書かれた議論の意味を表現する方法として、比較的安定的に抽出可能である言明間の類似度を使うこととした。自然言語で書かれた議論については、上記の議論グラフには類似度などがあるため、真・偽や真・偽・未定の論理値に基づく評価手法は使えない。連続値に基づく過去の研究としては、ある言明の値をその言明を攻撃する言明の値から何らかの関数により定める手法、言明とそれを攻撃する言明の間の値の連立方程式を生成しその解を値とする手法、何らかの確率分布を定め言明が真となる確率をその言明の値とする手法、議論グラフからゲームを構成しゲーム理論における報酬を値として使う手法、議論グラフをネットワークとみなしPageRank値を言明の値とする手法、議論グラフからニューラルネットワークを構成し学習したモデルの出力を値とする手法、議論グラフから生成した微分方程式の不動点を言明の値とする手法などがあるが、本研究の議論グラフに直接適用可能なものはなかった。そこで、マルコフ論理ネットワークの考えに基づいて、攻撃と支持、類似度、推定的証拠の原則を拘束条件として表現し、拘束条件の破れを最小化した時の議論要素の値を結果とすることを考えた。言明 a と b が両方成り立つことはないという拘束条件は、真を1、偽を0で表わした場合、 $ab = 0$ となるが、最小化で計算しやすい凸性を持たせるために $a^2b^2 = 0$ とした。また、 a が b を攻撃するという意味を考えると、 a が真の時は b が偽になって欲しいが逆は真ではない。そこで最小化の時に a は変更しないようにするために、 $a^2b^2 = 0$ 中の a を固定変数 \bar{a} として a と区別することにした。このようにして構成したコストの式から勾配法を用いると、自然に微分方程式が導けた。この微分方程式の解には不動点が複数ある場合やリミットサイクルになる場合がある。従来の評価手法では、値が一意に定まらないのは悪い性質と考えられてきたが、これらの場合は議論グラフの特徴を表していると考え、値を一つに定めたい時は、微分方程式の流れをマルコフ遷移と捉え、一様分布の初期値から始めた時の時刻 $t \rightarrow \infty$ での定常分布での期待値を値として使うことを考えた。実際の計算は初期値をサンプリングして行うことにすると、標本分散を用いた誤差の評価や階層クラスタリングによる値の分類などの考えに自然に到達した。分析手法については、コストの式に着目した分析手法を考案した。主張の拘束条件の不動点などでの値を拘束条件の破れの程度を表すものと考えられることができるが、主張に重みがかかっている場合は、重みを外したものを持ちいることで0 ~ 1の範囲の値に規格化でき、異なる議論グラフ間でも比較が可能となる。次に微分方程式を考えると、言明 a について、 $\frac{da}{dt} = -\frac{\partial \text{コスト}}{\partial a}$ なので、不動点では $\frac{\partial \text{コスト}}{\partial a} = 0$ である。そこでこの各項を釣り合いへの寄与として考えることができる。不動点である主張が値を決めるために働いているかは、固定変数以外の値を0から1に変えた時のその主張に対応するコスト中の項の値の変化(絶対値)で測ることができる。小さな議論グラフは過去の研究の例のものなどを利用した。大きな人工的議論グラフはICCAのものを使用した。また、自然言語で書かれた議論については、審議会の公開されている議事録やSNS上の公

開されている議論から協力者が作成した議論グラフを使用した。多人数での議論の議論グラフは様々な機会に協力者に書いてもらった議論グラフを使用した。

4. 研究成果

自然言語で書かれた議論を適切に表現できるような議論グラフのモデルを構築し、評価手法と分析手法を提案し、理論的性質を調べるとともに実際の議論グラフでその有効性を確かめた。

議論グラフは、言明(議論グラフのノードに対応)と主張(議論グラフのエッジに対応)からなり、自然言語で書かれた議論の攻撃や支持の関係を表現するためのものであるが、自然言語で記述された言明では、言明の間に攻撃や支持以外の関係が存在する。本研究では、そのような関係の中で比較的安定的に抽出できる言明間の類似度を対象とし、それを加えた議論グラフを定式化した。

定式化した議論グラフに対して、議論要素の成立の程度と評価手法を提案した。本研究では、攻撃と支持以外に類似度を扱うので、それらを含めて計算できる評価手法でなければならない。そのため、主張と類似度と推定的証拠の原則を、議論要素に対応する変数の式からなる制約条件と考えて、制約の破れの程度をコストとして表現し、コストの(主張の意図を実現するための制限を課した)最小化により、議論要素の値を求める手法を考案した。この最小化に勾配法を用いることで、時刻 t に関する微分方程式を得ることができる。この微分方程式の性質としては、1) 任意の初期値に対して解(時刻 $t \geq 0$ の値)が存在する。2) 値は $0 \sim 1$ の範囲にある。3) 不動点が1つは存在する。4) 不動点が複数ある場合やリミットサイクルがある場合がある。5) 議論要素の値が満たすべきものとして提案されている公理の多くを満たしているが、Cardinality Precedence(攻撃元の数が多い言明の値の方が小さい)のように(値を考慮に入れないなど公理の元となる原理が単純すぎて)満たさないものもある。6) 推定的証拠の原則の影響を強くすると議論要素の値は1に近づく。7) 値の依存関係がサイクルを持たなければ不動点は一意に定まり、時刻 $t \rightarrow \infty$ でその不動点に収束する。8) 値の依存関係が同じである議論要素は対応する初期値から計算すると同じ時刻には同じ値を持つ。9) 議論グラフに議論要素を追加しても、追加された議論要素に依存しない議論要素の値は変化しない。10) 適当な仮定の元で、議論要素に攻撃を追加すると値は減少する。11) 適当な仮定の元で、攻撃を追加した時の議論要素の値 a の変化は a に比例し、支持を追加したときは $1 - a$ に比例する。12) SPURIという議論分析フレームワークでは議論要素の値を真・偽・不定で与えているが、SPURIで真の議論要素の値は1、偽の議論要素の値は $\frac{\lambda}{1+\lambda}$ (λ は推定的証拠の原則の重み)以下となる。13) SPURIという議論分析フレームワークでは主張のマクロ構造として論というのが導入されているが、論の間関係は値への影響と対応している。例えば、反論(攻撃への支持の連鎖) $s_1 + s_2$ (言明 s_1 が s_2 を支持)、 $s_2 + s_3, \dots, s_m - a_i$ (s_m が a_i を攻撃)と立論(支持の連鎖) $a_1 + a_2, \dots, a_m + a$ の関係が無効化というが、立論が与える a の値への影響は反論が減らすことになる。

不動点が複数ある場合やリミットサイクルになる場合は、議論要素が単純な値とならない。その場合、値の分布が定常分布となるものを考え、その分布による期待値を値とする方法を考案した。定常分布の取り方には任意性があるが、不動点の場合は、その不動点に収束する初期値の体積、リミットサイクルの場合は、そのリミットサイクルに収束する初期値の体積をそれぞれの確率とする。リミットサイクルでは一周期での期待値を値とする。議論要素の値の依存関係が同型なら、全ての初期値に対する議論要素の値の多重集合は一致するので、期待値も一致する。実際の計算では、初期値をランダムにサンプリングして計算する。リミットサイクルの場合は一周期の平均値を近似計算する必要がある。ここで、サンプルを増やすと、サンプル数の平方根に反比例して正しい期待値に収束する。この収束は議論要素ごとであり、議論要素の数には依存しない。これを利用して、二つの議論要素の値が一致しているかを判定する方法を考案した。多数の議論要素の値の一致を調べる場合のために、議論要素の値の差の絶対値を距離とした階層クラスタを用いた簡便な手法も提案した。この手法では、十分なサンプルを用いると議論要素のクラスタを統合し終わったところで、クラスタ間距離が急激に上昇することを利用して、

複数人が独立に書いた議論グラフを統合する実験を行った。複数人が独立に議論グラフを書くと、本来同一の言明を異なる表現で表すことがある。その点を評価において考慮するために、コストに類似度の項を入れたが、実際に使う場合は言明間の類似度を与える必要がある。そのために、精度が高いと言われている深層学習により学習したモデルを用いて類似度を計算した。議論の言明間の正しい類似度を与えるモデルを学習するには、類似した言明と類似していない言明の大量のデータが必要となるが、そのようなデータは用意できないことが多いので、一般的な文章を用いた事前学習モデルを、実際の議論のデータから抽出した少量のデータを用いてファインチューニングする手法を用いて、精度が向上することを確認した。また、その類似度を使って求めた値では、類似している言明の値が近くなっていることを確認した。

分析手法として、残留コスト・釣り合い・有効度という分析指標を考案した。残留コストは不動点における主張ごとに制約の破れの程度を表すもので、制約が完全に満たされれば0、そうでなければ1の値を持つ。どの主張の残留コストも値の範囲が $0 \sim 1$ であるので、議論グラフが異なっても比較可能である。従って、議論が進み、議論グラフが拡大していくにつれて、同じ主張の残留コストがどのように変化するかを追跡することが可能である。釣り合いは不動点における主張、類似度、推定的証拠の原則の強さからなる。これは釣り合った状態でどの制約が支配的であるかを見るために使うことができる。有効度は不動点において、主張が値を決めるために有効に働いているかを見るためのもので、主張の制約が値をどの程度決めていくかを示すものである。

実験は、単純な議論グラフ、大きな人工的議論グラフ、審議会とSNSの議論グラフ、あるトピックに関する議論グラフに対して行った。単純な議論グラフでの検討から、複数の不動点がある場合やリミットサイクルがある場合があることがわかった。言明 a が b を攻撃し、 b が a を攻撃しているような場合は、 a が大きく、

b が小さい不動点 f_1 と、 a が小さく、 b が大きい不動点 f_2 、 a と b が中間的な値を取る不動点 f_3 の3つの不動点が生じる。 f_3 は不安定、 f_1, f_2 は安定な不動点である。 $a > b$ の初期値からは f_1 、 $a < b$ の初期値からは f_2 、 $a = b$ の初期値からは f_3 の不動点に収束する。これは最初の思い込みがそのまま固定されるものと解釈できる。なお、推定的証拠の原則の拘束条件を強くすると、 f_1, f_2, f_3 が合流し、どのような初期値からも a と b が中間的な値を取るようになる。言明 a_1 が a_2 を攻撃し、 a_2 が a_3 を攻撃し、...、 a_n が a_1 を攻撃するような攻撃の輪で、 n が偶数の場合は上記と同様に a_1 が大きい f_1 と a_1 が小さい f_2 と a_1 が中間的な値となる f_3 が生じる。 n が奇数の場合は f_1, f_2 のような安定状態はできず、リミットサイクルが生じる場合があることを発見した。

議論グラフの特徴を持つと期待される人工的に生成された議論グラフを集めたICCMAがある。ICCMAのいくつかの議論グラフについて値の計算をおこなった。その結果には、複数の不動点があるものがあるが、議論要素の値の大きさを初期値のサンプルごとにアイテム集合として相関分析を行い、大きなアイテム集合を取り出すと、議論要素 a の値が大きい場合は b の値は小さいというような傾向を抽出することができた。このように議論要素の値は一定の解釈のもとで大まかな傾向を抽出するために使うことができる。議論の進展に応じてこのような傾向を抽出するためにはシーケンシャルパターンマイニングが必要となる。そこで、シーケンシャルパターンマイニングによる分析の性格を知るために、制御しやすいデータであるチェスプレイヤーの棋譜を用いた分析を行い、分析の可能性を示した。

法制審議会における成人年齢引き下げの議論と第二東名建設についてのSNSでの議論の比較を行った。法制審議会の(最終回の1回前の)議論では、結論の値が1に近くなっており、結論の是非が決着している。それに対してSNSでの議論は結論の値が中間的に値になっており、結論の是非が決着していない。残留コストでも、法制審議会の議論では、残留コストが0に近くなっており、結論への反論は全て反論されて残されていない。これに対して、SNSでの議論は残留コストが残っているものが多くあり、言いつばなしの言明が多く残されていることがわかった。これは議論がどの程度収束しているかを判定するために残留コストを使う可能性を示唆するものである。

(フェイク情報に関連して)SNSでの投稿の制限の是非に関する議論では、自然言語で書いてもらった言明の類似度を、(一般的な文章を用いた)事前学習(議論の特徴で)ファインチューニングする深層学習を用いたモデルを使って計算した。その類似度を使った計算により、議論グラフの拘束の元で類似した言明の値を近づけることができた。類似度の残留コストが大きな2つの言明はそれらに対する攻撃や支持の構造が大きく異なっており、言明の見かけは似ているが、議論におけるポイントが異なるものであった。このように自然言語で書かれた言明を正しく解釈するためには、その背景となる知識も踏まえる必要がある。その方法を検討するために、背景知識がBOK(Body Of Knowledge)として整理されている計算機科学の代表的な教科書を、BOKに基づいたトピックモデルを用いて分析を行った。

倫理問題として有名なトロリー問題について、言明を与えた上で主張を複数人に書いてもらった議論グラフを検討した。この議論では、言明は共通であるが、どの言明間に主張を書くかは自由としたため同じ言明間に複数の主張が生じることがあるが、それらを(記述した人数比の)重みをつけて表すことで一つの議論グラフに統合できるように議論グラフの定式化を拡張した。また、その重みをコストの式でも用いることで、評価方法を統合した議論グラフに適用できるようにした。一人の議論グラフはサイクルを持たないが、統合した議論グラフにはサイクルが存在することがあり、不動点が一つであることは保証できないが、この実験で統合した議論グラフでは、一つの不動点に収束した。その時のサイクル中の主張を見ると、残留コストが大きな主張、有効度の小さい主張、残留コストが小さく有効度の高い主張の3つがあった。残留コストが大きな主張と有効度の小さい主張は議論グラフとの整合性が低く、残留コストが小さく有効度の高い主張は議論グラフ全体と整合性が高く、正しい主張と考えられる。発言順などの間接データと整合していることはこの判断の傍証といえる。これは他の部分と整合性の低い主張の発見のために残留コストと有効度を使う可能性を示唆するものである。また、結論は中間よりやや1に近い値となったが、釣り合いを調べると、1に近づけるもので主要な項は推定的証拠の原則であった。法制審議会の議論やSNSの議論では、1に近づける主要な項は支持であり、結論を積極的に支持する主張がされている。結論への積極的な支持がないことは、この倫理問題の性格を反映していると考えられ、釣り合いの利用の一つの可能性を示唆するものとなっている。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 山口和紀, 松田源立, 柴田裕介	4. 巻 64-8
2. 論文標題 議論分析フレームワークDECRA	5. 発行年 2023年
3. 雑誌名 情報処理学会論文誌	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 3件）

1. 発表者名 Kawamata Taisuke, Matsuda Yoshitatsu, Sekiya Takayuki, Yamaguchi Kazunori
2. 発表標題 Analysis of Computer Science Textbooks by Topic Modeling and Dynamic Time Warping
3. 学会等名 IEEE TALE（国際学会）
4. 発表年 2021年

1. 発表者名 Yamaguchi Kazunori, Matsuda Yoshitatsu
2. 発表標題 Cost-based Framework for Natural Language Argumentation Analysis
3. 学会等名 IEEE ITHET（国際学会）
4. 発表年 2021年

1. 発表者名 小林雅治, 松田源立, 川又泰介
2. 発表標題 シーケンシャルパターンマイニングを利用した棋譜に基づくチェスプレイヤーの棋風の分析
3. 学会等名 情報処理学会全国大会
4. 発表年 2020年～2021年

1. 発表者名 Yamaguchi Kazunori, Matsuda Yoshitatsu, Morinaga Yuya
2. 発表標題 Averaging Solution of Differential Equations as Value for Argumentation Graph with Similarity
3. 学会等名 Knowledge-Based and Intelligent Information & Engineering Systems (KES) (国際学会)
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	松田 源立 (Matsuda Yoshitatu) (40433700)	成蹊大学・理工学部・准教授 (32629)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------