

Graph Attention Networkによる Argumentationのコンポーネント分類

鈴木 祥太^{1,a)} 伊藤 孝行^{1,b)}

概要: 議論マイニングは Argumentation を解析し、その構造を特定することを目的とする。議論マイニングにおいて、コンポーネント分類は重要な課題である。コンポーネント分類を行うため、既存の手法は、複雑な議論構造をベクトルのような簡単な表現の特徴量に変換する。しかしながら、これらの特徴量に基づく手法では、複雑な構造を扱う上で貴重な情報が失われると考えられる。この問題を解決するため、本稿では、議論構造を直接的に学習することで、コンポーネント分類を行う手法を提案する。議論構造を直接的に学習するために、提案手法は Graph Attention Network を用いる。提案手法を評価するため、評論のコーパスを用いて実験を行った。実験の結果、提案手法は既存の特徴量に基づく手法よりも正確にコンポーネント分類を行うことが示された。

キーワード: 自然言語処理, 議論マイニング, 深層学習

Argument Component Classification with Graph Attention Network

Abstract: Argument mining aims to analyze argumentation and identify its argumentative structure. In the field of argument mining, argument component classification is an important task. To address this task, several studies have transformed complex argumentative structures into features of simpler representations such as vectors. However, these feature-based approaches are usually believed to lose valuable information for dealing with these complex structures. To tackle this problem, we propose an approach that performs argument component classification with directly learning argumentative structures. Towards this end, the proposed approach employs a graph attention network. To evaluate the proposed approach, we conducted a set of experiments on a corpus of persuasive essays. The experimental results show that the proposed approach performs argument component classification more accurately than the existing feature-based approaches.

Keywords: Natural Language Processing, Argument Mining, Deep Learning

1. はじめに

Argumentation は、何かを主張したり、質問したり、否定したり、これらの主張、質問、否定に答えたりすることを目的とした言語活動である [29]。また、Argumentation は、人々が意見を出し合い、自身の考えの長所と短所を考える社会活動である。Argumentation は、人と人とのコミュニケーションが目的を果たすことができるように、コ

ミュニケーションの原則を定める。特に、意見の相違を解決する際には、立場を認識することが非常に重要である。

議論マイニング (Argument Mining) は、評論 [18], [26], [27], 法的文書 [16], レビュー [11], オンライン議論 [7] などの自然言語のテキストから、Argumentation を解析し、その構造を特定することを目的とする研究分野である。議論マイニングは、さまざまな分野への応用の可能性が豊富であることから、大きな注目を集めている。議論マイニングは、人々が Argumentation を解析し、その構造を特定する方法に沿って、3つのタスクに分類される。一つ目のコンポーネントの識別 (Component Identification) では、コンポーネントであるテキストとコンポーネントでないテキス

¹ 名古屋工業大学
Nagoya Institute of Technology, Nagoya, Aichi 466-8555, Japan

^{a)} suzuki.shota@itolab.nitech.ac.jp

^{b)} ito.takayuki@nitech.ac.jp

トを分離する。二つ目のコンポーネント分類 (Component Classification) では、コンポーネントを Claim や Premise などの異なる種類に分類する。三つ目の議論構造の識別 (Structure Identification) では、コンポーネントをつなぎ、支持関係 (Support) と攻撃関係 (Attack) の議論関係を認識する。

コンポーネント分類を行うため、既存の手法は、複雑な議論構造をベクトルのような簡単な表現の特徴量に変換し、サポート・ベクタ・マシン [16], [26], [27], 系列ラベリング [7], ポインタ・ネットワーク [18] を訓練した。例えば、既存の手法 [7], [16], [18], [26], [27] は、語彙の、構造の、指標の、文脈の、構文の、確率の、談話の、埋め込みの特徴量を作成した。しかしながら、これらの特徴量に基づく手法では、複雑な議論構造を扱うには不十分であると考えられる [24]。特に、複雑な構造を簡単な表現の特徴量に変換することで、複雑な構造を扱う上で貴重な情報が失われる [6]。

この問題を解決するため、本稿では、既存の特徴量に基づく手法とは異なり、議論構造を直接的に学習することで、コンポーネント分類を行う手法を提案する。議論構造を直接的に学習するために、提案手法はグラフ・ニューラル・ネットワーク [6], [9], [24] の一種である Graph Attention Network [30] を用いる。グラフ・ニューラル・ネットワークは、ニューラル・ネットワークを用いてグラフ構造を直接的に符号化する。特に、Graph Attention Network は、注意機構に基づく構造のため、異なるグラフ構造を扱うことができる。Graph Attention Network を用いることで、提案手法はコンポーネントの数と議論関係の数の種類が豊富である自然言語のテキストを処理することができる。

提案手法を評価するため、評論のコーパス [27] を用いて実験を行った。このコーパスはユビキタス知識処理研究所 (Ubiquitous Knowledge Processing Lab) が公開しており*1, 402 編の評論から構成される。また、このコーパスは、各コンポーネントについて詳細に説明するガイドラインが定める基準でアノテーションされる。アノテーションの結果、Fleiss のカッパ係数 [4] と Krippendorff のアルファ係数 [14] の値は、人間のアノテータの間で相当の合意を示す。実験の結果、提案手法は既存の特徴量に基づく手法 [18], [27] よりも正確にコンポーネント分類を行うことが示された。

本稿の以降の章は、次のように構成される。2 章では、コンポーネント分類と Graph Attention Network の分野における関連研究について議論する。3 章では、議論構造を直接的に学習することでコンポーネント分類を行う提案手法を詳述する。4 章では、提案手法を評価するために行った実験の設定を説明する。5 章では、実験結果を詳細に説

明する。6 章では、実験結果について考察する。7 章では、本稿をまとめ、今後の展望を述べる。

2. 関連研究

2.1 コンポーネント分類

コンポーネント分類は、コンポーネントを Claim や Premise などの異なる種類に分類することを目的とする [27]。Palau と Moens の研究 [16] は、コンポーネントを Premise と Conclusion に分類することを目的とした。コンポーネントを Premise と Conclusion に分類するために、彼らの手法は、最大エントロピー分類器とサポート・ベクタ・マシンを採用した。また、Stab と Gurevych の研究 [26] は、コンポーネントを Major Claim, Claim, Premise, Non-Argumentative に分類することを目指した。コンポーネントをこれら 4 種類に分類するために、彼らの手法は、評論からさまざまな種類の特徴量を作成して、サポート・ベクタ・マシンを採用した。さらに、Habernal と Gurevych の研究 [7] は、コンポーネントを Claim, Premise, Backing, Rebuttal, Refutation に分類することを目指した。コンポーネントをこれら 5 種類に分類するために、彼らの手法は、コンポーネントの識別と分類を系列ラベリングにより行った。その上、Potash らの研究 [18] は、コンポーネントを Major Claim, Claim, Premise に分類することを目的とした。コンポーネントをこれら 3 種類に分類するために、彼らの手法は、ポインタ・ネットワークを採用した。

既存の手法はいずれも、自然言語のテキストからさまざまな種類の特徴量を作成して、最大エントロピー分類器、サポート・ベクタ・マシン、系列ラベリング、ポインタ・ネットワークといった特徴量に基づく分類器を採用した。

2.2 グラフ・ニューラル・ネットワーク

グラフ・ニューラル・ネットワーク [6], [24] は、ニューラル・ネットワークを用いてグラフ構造を直接的に符号化する。グラフ・ニューラル・ネットワークは、ニューラル・ネットワークが医療画像診断、分子生物学、化学、音声処理、自然言語処理などの構造化されたデータを扱うことができることを示した。Sperduti と Starita の研究 [24] に端を発する。特に、彼らの研究は、標準的なニューラル・ネットワークのような特徴量に基づく手法では、リスト、木、グラフといったその大きさや複雑さが変化する複雑な構造を扱うには不十分であると考えられていると指摘した。これらの複雑な構造を処理するため、Gori らの研究 [6] は、特徴量に基づく手法に対立するものとして、グラフを直接的に処理することができるニューラル・ネットワークであるグラフ・ニューラル・ネットワークを提案した。彼らの研究は、グラフ・ニューラル・ネットワークを有向グラフ、無向グラフ、ラベル付きグラフ、環状グラフに適用することができることを示した。特に、従来の機械学習手法では、

*1 www.ukp.tu-darmstadt.de/data/argumentation-mining

複雑な構造をベクトルのような簡単な表現の特徴量に変換することによって、価値のある情報が失われてしまうと指摘した。貴重な情報が欠如のために、従来の特徴量に基づく手法は、汎化性能の低下に陥ると論じた。その後、多くの分野において、グラフ・ニューラル・ネットワークに基づく手法であるリカレント・グラフ・ニューラル・ネットワーク [21]、コンボリユーションアル・グラフ・ニューラル・ネットワーク [2], [10]、変分グラフ・オート・エンコーダ [13]、時空間グラフ・ニューラル・ネットワーク [23] が提案された。これらのグラフ・ニューラル・ネットワークに基づく手法は、グラフ・ニューラル・ネットワークが多くの種類の複雑な構造を学習することを実証した。

2.3 Graph Attention Network

Graph Attention Network[30] は、グラフ・ニューラル・ネットワークに基づく手法であり、グラフ構造を直接的に符号化する。さらに、Graph Attention Network は、注意機構に基づく構造のため、異なるグラフ構造を処理することができる。一方、通常のグラフ・ニューラル・ネットワークは、その学習済フィルタがグラフ・ラプラシアン固有値に依存するため、異なるグラフ構造を扱うことができない [12]。異なるグラフ構造を扱うことができる利点を持つグラフ・ニューラル・ネットワークは、Cora, Citeseer, Pubmed[22], Protein-Protein Interaction[8] といった4種類のグラフに基づくデータセットで最高水準の性能を達成した。この結果により、Graph Attention Network は、近傍全体を観測することで高い汎化性能を持つことが示された。

Velivckovic ら [30] の手法は、グラフ・ニューラル・ネットワークを引用関係ネットワークと生化学の分野に適用していたが、提案手法は、議論マイニングの分野に適用する。故に、提案手法が Graph Attention Network を用いてコンポーネント分類を行うことは独創性に富む。

3. 提案手法

3.1 提案手法の概要

提案手法の概要を図1に示す。提案手法は、評論をグラフに変換する部分と、変換したグラフの構造を直接的に学習してコンポーネント分類を行う部分で構成される。

3.2 評論からグラフへの変換

提案手法は、評論を、そのコンポーネントと議論関係から構成されるグラフに変換する。評論におけるコンポーネントと議論関係がそれぞれ、グラフにおけるノードとリンクに相当する。議論関係について、支持関係 (Support) と攻撃関係 (Attack) のように種類の異なる関係が存在する可能性があるが、どのような種類の議論関係も等しくリンクとする。また、グラフにおけるリンクには、各ノードから

そのノード自身への自己ループも含まれる。

3.3 グラフ構造の直接的学習によるコンポーネント分類

提案手法は、変換したグラフの各ノードのテキストをベクトル空間へ埋め込み、特徴量 $\mathbf{h} = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_N\}$, $\vec{h}_i \in \mathbb{R}^F$ (N はノードの数, F はベクトルの次元数) を構築する。これらの特徴量は、第一の Graph Attention Network 層の入力となる。Graph Attention Network 層は、各ベクトルを、ノードのテキストの埋め込み表現とリンクに基づき更新したベクトルの集合 $\mathbf{h}' = \{\vec{h}'_1, \vec{h}'_2, \dots, \vec{h}'_N\}$, $\vec{h}'_i \in \mathbb{R}^{F'}$ (F' はベクトルの次元数) を出力する。特に、特徴量 \vec{h}_i から \vec{h}'_i への更新では、ノード i を含むノード i の隣接ノードの特徴量 \vec{h}_j の線形結合を算出し、各注意機構の出力を結合する。

$$\vec{h}'_i = \left\| \sum_{k=1}^K \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^k \mathbf{W}^k \vec{h}_j \right) \right. \quad (1)$$

ここで、 $\|$ はベクトルの結合、 σ は活性化関数、 \mathcal{N}_i はノード i の隣接ノード、 α_{ij}^k は k 番目の注意機能、 $\mathbf{W}^k \in \mathbb{R}^{F' \times F}$ は重み行列を表す。 α_{ij} は、次式により算出される。

$$\alpha_{ij} = \frac{\exp(a(\mathbf{W}\vec{h}_i, \mathbf{W}\vec{h}_j))}{\sum_{k \in \mathcal{N}_i} \exp(a(\mathbf{W}\vec{h}_i, \mathbf{W}\vec{h}_k))} \quad (2)$$

ここで、 $a: \mathbb{R}^{F'} \times \mathbb{R}^{F'} \rightarrow \mathbb{R}$ は、ノード i にとってのノード j の重要度を表す注意機構である。また、第二の Graph Attention Network 層は、次式のように、結合の代わりに平均を算出し、ソフトマックス関数 [5] による出力の正規化を行う。

$$\vec{h}'_i = \text{softmax} \left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in \mathcal{N}_i} \alpha_{ij}^k \mathbf{W}^k \vec{h}_j \right) \quad (3)$$

第二の Graph Attention Network 層の各出力 \vec{h}'_i に対し、Major Claim, Claim, Premise のうち最も高い確率を示すクラスを、そのノードの予測とする。正解ラベルは、ノードが実際に該当するクラスである。提案手法は、教師あり学習を行い、交差エントロピーの損失値 \mathcal{L} を最小化することで、隠れ状態の重みを調整する。

$$\mathcal{L} = - \sum_{i \in N} \sum_{c \in C} t_{ic} \ln h'_{ic} \quad (4)$$

ここで、 C は {Major Claim, Claim, Premise} のようなコンポーネントのクラスの集合、 t_{ic} はノード i が実際にクラス c に該当する場合は1、該当しない場合は0、 h'_{ic} はノード i がクラス c に該当する予測確率である。

4. 実験設定

4.1 データセット

提案手法を評価するために実験を行った。実験では、評論のコーパス [27] を用いた。このコーパスは、402編の英

コンポーネントのテキスト	識別子	議論関係
we should attach more importance to cooperation during primary education	MC_1	$Cl_1 \rightarrow Pr_1$
through cooperation, children can learn about interpersonal skills which are significant in the future life of all students	Cl_1	$Cl_1 \rightarrow Pr_2$ $Cl_1 \rightarrow Pr_3$
What we acquired from team work is not only how to achieve the same goal with others but more importantly, how to get along with others	Pr_1	$Cl_2 \rightarrow Pr_4$
During the process of cooperation, children can learn about how to listen to opinions of others, how to communicate with others, how to think comprehensively, and even how to compromise with other team members when conflicts occurred	Pr_2	$Cl_3 \rightarrow Pr_5$
All of these skills help them to get on well with other people and will benefit them for the whole life	Pr_3	$Cl_3 \rightarrow Pr_6$
the significance of competition is that how to become more excellence to gain the victory	Pr_4	
competition makes the society more effective	Cl_2	
when we consider about the question that how to win the game, we always find that we need the cooperation	Pr_5	
Take Olympic games which is a form of competition for instance, it is hard to imagine how an athlete could win the game without the training of his or her coach, and the help of other professional staffs such as the people who take care of his diet, and those who are in charge of the medical care	Pr_6	
without the cooperation, there would be no victory of competition	Cl_3	
a more cooperative attitudes towards life is more profitable in one's success	MC_2	

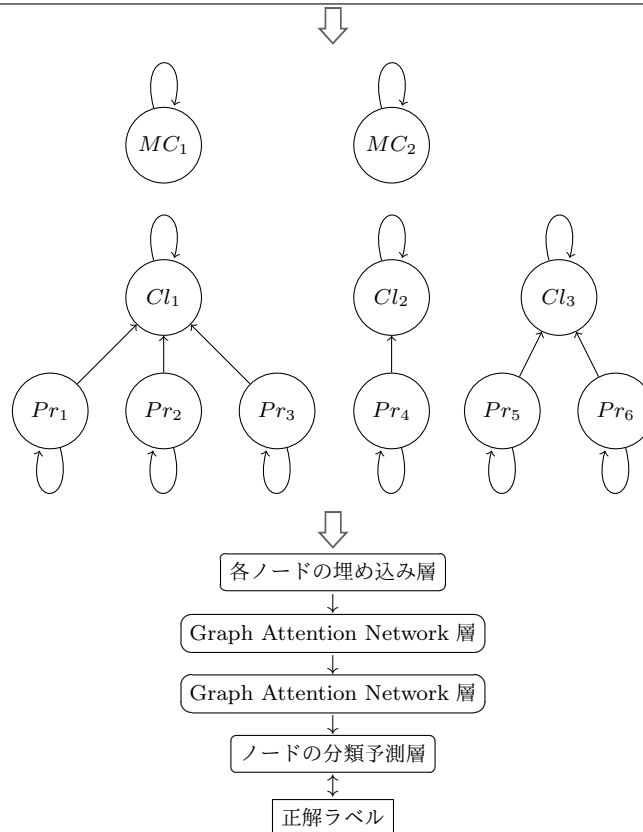


図 1 提案手法の概要

語評論を含み、1833 段落、7116 文、147,721 トークンから構成される。402 編のうち、訓練データとして 258 編、検証データとして 64 編、テストデータとして 80 編の評論を使用した。このコーパスにおいて、コンポーネントは Major Claim, Claim, Premise に分類される。また、アノテータ

のためのガイドラインでは、各コンポーネントのアノテーションの基準について詳細に記述される。

- Major Claim は、評論の話題に対する著者の姿勢を表す。通常、Major Claim は、評論の序論、もしくは結論、あるいはその両方に表れる。序論において、Major

Claim は話題に対する一般的な主張や意見であり、結論において、Major Claim は著者の姿勢における論旨の要約であるという特徴を持つ。

- Claim は、著者の姿勢を直接的に支持したり、反対したりする。つまり、Claim は、Major Claim を支持したり、反対したりする理由を述べたものである。
- Premise は、Claim を支持したり、反対したりする。すなわち、Premise は、Claim の真偽を読み手に納得させるための正当化や反論と考えられる。Claim のアノテーションとは対照的に、Premise は、その姿勢を属性に含まない。また、Premise と Claim や、その他の Premise とを結ぶ関係は支持関係 (Support) か攻撃関係 (Attack) とアノテーションされる。

コーパスとガイドラインを、ユビキタス知識処理研究所が公開している。また、コーパスはガイドラインが定める基準でアノテーションされ、Fleiss のカッパ係数 [4] と Krippendorff のアルファ係数 [14] の値は、人間のアノテータの間で相当の合意を示す。従って、コーパスは、提案手法の評価に適している。

4.2 ベースライン

提案手法を、コンポーネント分類を目的とする既存の特徴量に基づく手法と比較する。特に、最高水準の性能を達成した、サポート・ベクタ・マシンに基づく手法 [27]、ポイント・ネットワークに基づく手法 [18] を比較対象とする。

4.3 モデル設定

提案手法を実装するために、オープンソースの機械学習フレームワークである PyTorch [17] を使用した。また、PyTorch と親和性が高く、グラフ・ニューラル・ネットワークを簡単に実装することができるように構築された Python パッケージである、ディープ・グラフ・ライブラリ [31] を用いた。Graph Attention Network 層の隠れ状態の次元数を 256、Graph Attention Network 層のドロップアウト [25] の割合を 0.1 とした。また、注意機構の隠れ状態の次元数を 4、注意機構の出力の次元数を 6、注意機構のドロップアウトの割合を 0.1 とした。

学習は、バッチサイズを 80 とするミニバッチ学習を使用した。提案手法のパラメータの最適化には確率的勾配降下法 [20] を使用し、学習率を 0.01、モメンタムを 0.9 とした。さらに、学習率を調整する手法である Stochastic Gradient Descent with Warm Restarts (SGDR) [15] を使用し、300 エポックから SGDR による学習率の調整を開始した。学習は最大 4000 エポック行い、5 エポックごとに算出したスコアが最も高いモデルにより評価した。

4.4 特徴量

Stab らの研究 [27] を参考に、第一の Graph Attention

Network 層の入力となる以下の特徴量を実装した。

- 語彙の特徴量：見出語化されたユニグラムと、出現頻度が上位 2000 以上の単語のペアから構成される。
- 構造の特徴量：評論内でのコンポーネントの位置と、コンポーネントの統計量から構成される。
- 指標の特徴量：前方指標、後方指標、論的指標、反証指標、一人称を表す指標を含むか否か示す。
- 文脈の特徴量：前方指標、後方指標、論的指標、反証指標がコンポーネントに続くか否か示す。
- 構文の特徴量：コンポーネントの品詞タグ付け [1]、コンポーネントを含む文の従属節の数、構文木の深さ、主動詞の時制、助動詞の有無から構成される。
- 埋め込みの特徴量：BERT [3] による単語のベクトル表現から構成される。

4.5 評価尺度

評価尺度として F 値 [19], [28] を用いる。F 値は、適合率と再現率の調和平均である。

$$F \text{ 値} = \frac{2}{\frac{1}{\text{適合率}} + \frac{1}{\text{再現率}}} \quad (5)$$

適合率は、各クラスに該当すると予測したコンポーネントのうち、実際にそのクラスに該当するコンポーネントの割合である。再現率は、実際に各クラスに該当するコンポーネントのうち、そのクラスに該当すると予測したコンポーネントの割合である。各コンポーネントの分類性能と総合的な分類性能を評価するため、Major Claim, Claim, Premise の F 値とそれらのマクロ平均を算出する。

5. 評価結果

表 1 に、各モデルの評価結果を示す。結果から、提案手法はポイント・ネットワークに基づく手法より、Claim の F 値で 13.2%、Premise の F 値で 4.1% 上回る性能を示した。Graph Attention Network を用いて議論関係を直接的に学習することで、隣接ノードを持つ Claim と Premise のノードをより正確に分類できる。特に Claim の分類について、Sperduti と Starita の研究 [24] と Gori らの研究 [6] が指摘するように、ベースラインでは議論構造を簡単な表現の特徴量に変換することによって価値のある情報が失われてしまう一方、提案手法では Graph Attention Network を用いて議論構造を直接的に学習することによって、性能が向上した。一方、Major Claim の F 値は、ベースラインの F 値より低い値を示した。Major Claim のノードは、その隣接ノードを持たないため、議論関係の直接的学習による性能向上が見られない。また、提案手法はベースラインである特徴量に基づく手法より、F 値のマクロ平均で 4.8% 上回った。議論関係の直接的学習により、コンポーネント分類を行う総合的な性能の向上を示した。

表 1 提案手法と各ベースラインの F 値

手法	Major Claim	Claim	Premise	F 値のマクロ平均
提案手法	0.865	0.864	0.962	0.897
サポート・ベクタ・マシン [27]	0.891	0.682	0.903	0.826
ポインタ・ネットワーク [18]	0.894	0.732	0.921	0.849

表 2 提案手法による各コンポーネントの適合率, 再現率, F 値

コンポーネント	適合率	再現率	F 値
Major Claim	0.849	0.882	0.865
Claim	0.872	0.855	0.864
Premise	0.962	0.962	0.962

Ground Truth	Prediction		
	Major Claim	Claim	Premise
Major Claim	135	14	4
Claim	17	260	27
Premise	7	24	778

図 2 混同行列

6. 考察

6.1 各コンポーネントに対する分類性能の分析

提案手法の詳細な分類性能を分析するために, 表 2 に, 各コンポーネントの分類について適合率, 再現率, F 値を示す. さらに, 図 2 に, コンポーネント分類結果に関する混同行列を示す. 表 2 から, Major Claim について再現率が適合率より高い値を, Claim については適合率が再現率より高い値を示した. それぞれの結果と, 図 2 の混同行列から, 実際に Major Claim に該当するコンポーネントを正しく予測できていること, 実際に Premise に該当するコンポーネントを正しく予測できていることを示唆する. 一方, 実際は Claim に該当するコンポーネントの一部を Major Claim や Premise に該当すると予測してしまうことは, さらなる改善の余地があることが伺える. 更なる性能向上のためには, コンポーネントを Claim, Premise, Backing, Rebuttal, Refutation に分類している Habernal と Gurevych の研究 [7], Major Claim, Claim, Premise に分類している Stab と Gurevych の研究 [27], Potash らの研究 [18] を参考に, 実際に Claim に該当するコンポーネントを正しく分類するために有効な特徴量を調査する必要がある.

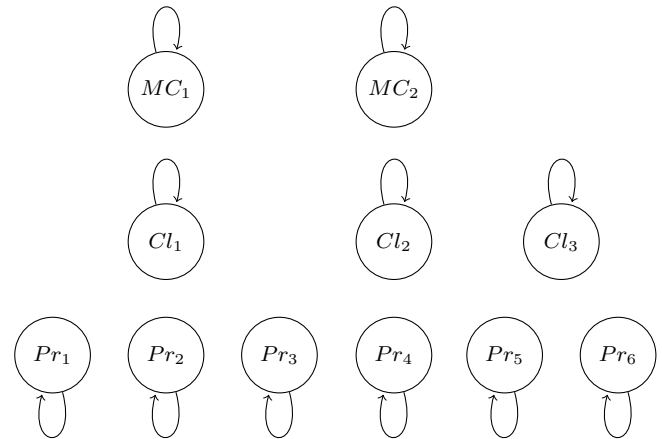


図 3 議論関係を含まないグラフ

6.2 学習した議論関係を分類に用いる有効性の分析

提案手法における, 学習した議論関係を分類に用いる有効性を検証するために, 図 3 に示すような議論関係を含まないグラフに対する分類性能の分析を行った. 分析では, 4 章の実験で得られた学習済モデルを使用した. 表 3 に, 議論関係を含まないグラフにおけるコンポーネント分類の評価結果を示す. 結果の通り, Major Claim, Claim, Premise といった全ての分類において, 議論関係を含むグラフに対す F 値は, 議論関係を含まないグラフに対す F 値より高い値を示した. 特に Claim の分類について, 議論関係を含むグラフに対する分類の F 値は, 議論関係を含まないグラフに対する分類の F 値より大幅に高い値を示した. この結果は, 学習した議論関係を分類に用いることが有効であることを示唆する.

7. おわりに

本稿では, 議論関係を直接的に学習することで, コンポーネント分類を行う手法を提案した. 提案手法は, 議論関係を直接的に学習するために, Graph Attention Network を用いる. 提案手法を評価するために, 評論のコーパスを用いて実験を行った. 実験の結果, 提案手法は, 既存の特徴量に基づく手法より正確にコンポーネント分類を行うことが示された. 今後は, 議論構造の直接的学習による議論構造の識別を行う.

謝辞 研究内容は, JST CREST 「エージェント技術に基づく大規模合意形成支援システムの創成: 代表伊藤孝行」(Grant 番号 JPMJCR15E1) に支援を受けている研究の一部である.

表 3 議論関係を含むグラフと議論関係を含まないグラフにおけるコンポーネント分類の F 値

グラフ	Major Claim	Claim	Premise	F 値のマクロ平均
議論関係を含む	0.865	0.864	0.962	0.897
議論関係を含まない	0.784	0.276	0.876	0.645

参考文献

[1] Brill, E.: Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging, *Computational linguistics*, Vol. 21, No. 4, pp. 543–565 (1995).

[2] Bruna, J., Zaremba, W., Szlam, A. and Lecun, Y.: Spectral networks and locally connected networks on graphs, *International Conference on Learning Representations (ICLR2014), CBLIS, April 2014* (2014).

[3] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* (2018).

[4] Fleiss, J. L.: Measuring nominal scale agreement among many raters., *Psychological bulletin*, Vol. 76, No. 5, p. 378 (1971).

[5] Goodfellow, I., Bengio, Y. and Courville, A.: *Deep learning*, MIT press (2016).

[6] Gori, M., Monfardini, G. and Scarselli, F.: A new model for learning in graph domains, *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, Vol. 2, IEEE, pp. 729–734 (2005).

[7] Habernal, I. and Gurevych, I.: Argumentation mining in user-generated web discourse, *Computational Linguistics*, Vol. 43, No. 1, pp. 125–179 (2017).

[8] Hamilton, W., Ying, Z. and Leskovec, J.: Inductive representation learning on large graphs, *Advances in neural information processing systems*, pp. 1024–1034 (2017).

[9] Hamilton, W. L., Ying, R. and Leskovec, J.: Representation learning on graphs: Methods and applications, *arXiv preprint arXiv:1709.05584* (2017).

[10] Henaff, M., Bruna, J. and LeCun, Y.: Deep convolutional networks on graph-structured data, *arXiv preprint arXiv:1506.05163* (2015).

[11] Kim, Y.: Convolutional Neural Networks for Sentence Classification, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1746–1751 (2014).

[12] Kipf, T. N. and Welling, M.: Semi-supervised classification with graph convolutional networks, *arXiv preprint arXiv:1609.02907* (2016).

[13] Kipf, T. N. and Welling, M.: Variational Graph Auto-Encoders, *stat*, Vol. 1050, p. 21 (2016).

[14] Krippendorff, K.: Measuring the reliability of qualitative text analysis data, *Quality and Quantity*, Vol. 38, pp. 787–800 (2004).

[15] Loshchilov, I. and Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts, *arXiv preprint arXiv:1608.03983* (2016).

[16] Palau, R. M. and Moens, M.-F.: Argumentation mining: the detection, classification and structure of arguments in text, *Proceedings of the 12th international conference on artificial intelligence and law*, ACM, pp. 98–107 (2009).

[17] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, *Advances in Neural Information Processing Systems 32* (Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. and Garnett, R., eds.), Curran Associates, Inc., pp. 8024–8035 (online), available from <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf> (2019).

[18] Potash, P., Romanov, A. and Rumshisky, A.: Here’s My Point: Joint Pointer Architecture for Argument Mining, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 1364–1373 (2017).

[19] Powers, D. M.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation (2011).

[20] Robbins, H. and Monro, S.: A stochastic approximation method, *The annals of mathematical statistics*, pp. 400–407 (1951).

[21] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M. and Monfardini, G.: The graph neural network model, *IEEE Transactions on Neural Networks*, Vol. 20, No. 1, pp. 61–80 (2008).

[22] Sen, P., Namata, G., Bilgic, M., Getoor, L., Galligher, B. and Eliassi-Rad, T.: Collective classification in network data, *AI magazine*, Vol. 29, No. 3, pp. 93–93 (2008).

[23] Seo, Y., Defferrard, M., Vandergheynst, P. and Bresson, X.: Structured sequence modeling with graph convolutional recurrent networks, *International Conference on Neural Information Processing*, Springer, pp. 362–373 (2018).

[24] Sperduti, A. and Starita, A.: Supervised neural networks for the classification of structures, *IEEE Transactions on Neural Networks*, Vol. 8, No. 3, pp. 714–735 (1997).

[25] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting, *The journal of machine learning research*, Vol. 15, No. 1, pp. 1929–1958 (2014).

[26] Stab, C. and Gurevych, I.: Identifying argumentative discourse structures in persuasive essays, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 46–56 (2014).

[27] Stab, C. and Gurevych, I.: Parsing argumentation structures in persuasive essays, *Computational Linguistics*, Vol. 43, No. 3, pp. 619–659 (2017).

[28] Ting, K. M.: Precision and Recall., *Encyclopedia of machine learning*, Vol. 781 (2010).

[29] van Eemeren, F. H., Grootendorst, R., Henkemans, F. S., Blair, J. A., Johnson, R. H., Krabbe, E. C., Plantin, C., Walton, D. N., Willard, C. A. et al.: Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments. (1996).

[30] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P. and Bengio, Y.: Graph attention networks, *arXiv preprint arXiv:1710.10903* (2017).

[31] Wang, M., Yu, L., Zheng, D., Gan, Q., Gai, Y., Ye, Z., Li, M., Zhou, J., Huang, Q., Ma, C., Huang, Z., Guo,

Q., Zhang, H., Lin, H., Zhao, J., Li, J., Smola, A. J. and Zhang, Z.: Deep Graph Library: Towards Efficient and Scalable Deep Learning on Graphs, *ICLR Workshop on Representation Learning on Graphs and Manifolds*, (online), available from (<https://arxiv.org/abs/1909.01315>) (2019).