

グラフ表現による文章の構造情報を用いた機械読解

小坂 直輝¹ 小林 哲則¹ 林 良彦¹¹ 早稲田大学理工学術院

kosaka@pcl.cs.waseda.ac.jp

概要

本研究では、グラフ構造による文章表現を利用する機械読解手法を提案する。単語依存構造をベースに文の位置関係や共参照、談話修辭構造を用いて文章全体の構造を表すグラフを生成し、これをグラフ畳み込みモデルを介して表現する。得られた表現を事前学習済言語モデルによる文脈表現と併用することで読解問題を解く。多肢選択型機械読解タスクにおける実験の結果、言語モデル単体を用いた場合や既存手法と比べて精度向上が確認され、文章全体の構造をグラフとして表現し利用することの有用性が示唆された。

1 はじめに

近年、BERT [1] や XLNet [2] といった大規模な事前学習済言語モデルが幅広いタスクで高い精度を達成しているが、こうしたモデルは語順や構文情報に弱いという指摘が存在する [3]。

多肢選択型機械読解においても言語モデルと併用して、文章の構造情報を明示的に与えることで精度向上を実現した研究が現れている ([4], [5])。一方でこれらの研究では、いずれも用いられている構造情報は限定的であるという問題があり、特に文間の関係や文章全体の構造などはあまり考慮されていない。文章構造の適切なグラフ表現形式を明らかにすることを目的とした我々の以前の研究 [6] においても、上記のような上位階層の構造情報は考慮できていなかった。

本研究では、グラフ構造による文章表現を利用する機械読解手法を提案する。より具体的には、単語依存構造をベースに文の位置関係や共参照、談話修辭構造を用いて文章全体の構造を表すグラフを生成し、これをグラフ畳み込みモデルを介して表現した結果を事前学習済言語モデルによる文脈表現と統合することにより読解問題を解く。多肢選択型機械読解のデータセット RACE[7] を用いた実験の結果、言

語モデル (BERT) 単体や文章の構造情報を部分的に利用する既存手法と比べて精度向上が確認され、提案する文章グラフを用いて文章全体の構造情報を明示的に与えることが有用であることが示唆された。

2 関連研究

文章の構造情報を事前学習済言語モデルと併用する形で利用し読解タスクに取り組んだ研究 [4] や [8] では、self-attention network 構造をもつエンコーダで文章を処理する際に、単語依存構造によりアテンション計算を明示的に制限してエンコードした表現を併用することで精度向上を達成している。本研究でも単語依存構造グラフを用いるという点で関連があるが、単語依存関係のみでは文間の意味関係や文章の談話構造といった高い階層の構造情報は考慮できない。加えて上記研究のようなアテンション計算に制約を加えるような手法には、依存関係ラベルなどの関係の種別を明示的に利用することはできないという問題もある。さらに高度な意味情報を用いた研究としては、意味役割付与を行った結果を入力する文の単語表現に加える形で用いて読解精度向上を達成した研究 [5] がある。

文章の構造情報をグラフとして表現し読解タスクに取り組んだ研究としては [9] がある。文章から文内もしくは文間にまたがる因果関係や時間的關係を表すイベントグラフと呼ばれるグラフを生成し、それを既存の機械読解モデルの補助として用いることで精度向上を達成している。また、[10] では抽象的意味表現 (Abstract Meaning Representation: AMR)[11] を用いて、機械読解における文章と QA の文それぞれから生成したグラフのアライメント計算を行うことで読解問題を解いている。これらの研究では文内に限定されない構造情報を表現したグラフを生成しているが、高い階層の構造情報の考慮は限定的である。また、両研究では事前学習済言語モデルとの比較や併用は行われていない。

文章のグラフ表現を文間関係抽出のタスクに適用

した研究例として [12] がある。この研究では、文の単語依存構造グラフに対して隣接語や隣接文を繋いだエッジ、共参照解決によるエッジなどを付加することにより、エッジに富んだ文章のグラフ表現を生成し、グラフ畳み込みモデルを介し利用している。こうしたエッジの追加は機械読解においても有用だと考えられ、本研究でも [12] と同様に、文の隣接関係や共参照解決結果を用いたエッジの追加を行う。

以上のように、機械読解をはじめ自然言語タスクにおいて文章の構造情報を明示的に与える研究は複数存在するが、総じて利用されている構造情報は限定的であり、特に文間の関係や文章全体の構造など上位階層の情報に関しては検討の余地が残されている。

3 アプローチ

本研究では、読解において対象となる文章全体からその構造を表す図 1 のようなグラフを生成し、得られた文章のグラフ表現をグラフ畳み込みモデルを介して、言語モデルと併用する形で利用することで読解問題を解く。以下では文章グラフ生成法と構築した読解モデルについて述べる。

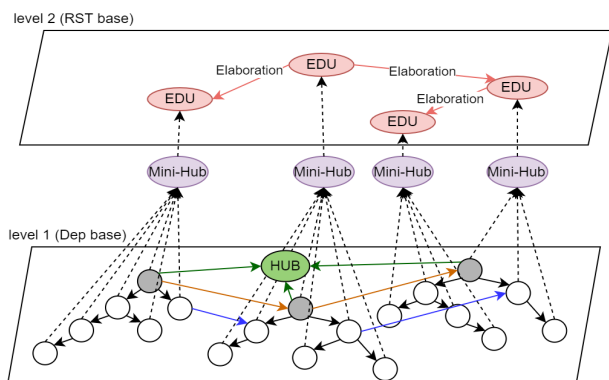


図 1 階層的な文章グラフの概要。level 1 では各文の依存構造を表し、level 2 では文章全体の修辞構造を表す。

3.1 文章のグラフ表現生成法

読解問題における文章のグラフ表現の生成においては、まず文章中の各文をグラフに変換し、文間にまたがるエッジを追加することで各文ごとのグラフを統合する。さらに談話修辞構造を利用することで文章全体の構造を表すグラフを構成する。

3.1.1 文グラフの生成と統合

ベースとなる各文の表現には、[4] や [8] と同様に依存構造グラフを用いる。文分割や依存構造解析に

は自然言語解析エンジンの spaCy¹⁾を用い、単語をノードとしその依存関係をエッジで表現することにより文の構文・意味構造を表す依存構造グラフを作成する。

各文のグラフの統合では、(1) 隣り合う文同士の依存関係グラフの ROOT ノードを隣接文エッジにより結ぶ、(2) 文統合の目的で特別に導入する HUB ノードを介して全ての文をスター状に繋ぐ、(3) 共参照関係にある単語間を結ぶ共参照エッジにより文グラフを補強する、という 3 つの手段 [6] を適用する。ここで共参照解決エンジンには neuralcoref²⁾ [13] を用いた。

以上までの段階で生成される文章のグラフ表現の例を図 2 に示す。

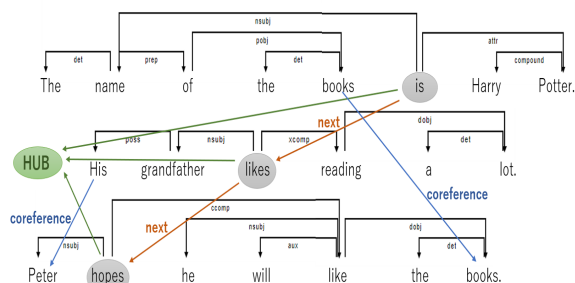


図 2 単語依存構造グラフとその統合

3.1.2 RST を利用した文章グラフの生成

図 2 のグラフでは文間の関係や文章全体の構造を考慮しきれていないとは言えない。そこでより高い階層の構造情報を取り入れるために、文章の部分間の関係を記述するテキスト編成の理論である修辞構造理論 (Rhetorical Structure Theory: RST) [14] に基づいて構成される修辞構造木を利用する。

修辞構造木において、文章は Elementary Discourse Unit (EDU) と呼ばれる基本談話単位に分割され、隣接する談話単位同士は修辞関係によってより大きい談話単位へと統合されていく。修辞関係には ‘Elaboration’ や ‘Condition’ など、その関係を表す様々なラベルが与えられ、その中には因果関係や時間的な関係を表すものも含まれる。こうした情報は読解を行う際の推論に有用だと考えられる。

図 3 は読解問題の文章の一部を修辞構造木に変換した例である。図 3 では There are two pictures in the room, too. という文を They are on the wall. という文が精緻化しているという関係が表現されている。こ

1) <https://spacy.io/>

2) <https://github.com/huggingface/neuralcoref>

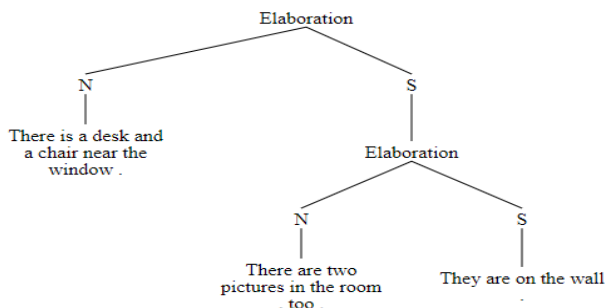


図3 RSTに基づく修辞構造木

の文章に対しては部屋の壁には何があるかという質問が存在し、こうした文中の特定の一文のみを見ても答えるのが難しい読解問題の推論においてRSTの修辞構造は有用だと考えられる。

実際のグラフ生成においては、基本談話単位であるEDUのノード同士を修辞関係を表すエッジで直接結ぶことによりグラフを構築する。RSTパーサーには[15]を用いた。例えば、図3のような修辞構造木は、図4のようなグラフとして表される。

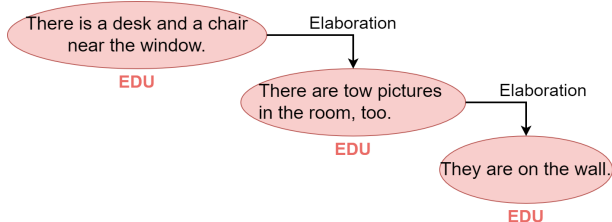


図4 RST グラフ

3.1.3 異なる階層の構造情報の統合

談話修辞構造の利用法としては、単語依存構造をベースとしたグラフ(図2)に、さらにエッジを追加するような形で一つのグラフに統合する手法がまず考えられる。しかし、単語依存構造と談話修辞構造はレベルの大きく異なる別々の構造情報であることから、二種類の構造情報を混ぜてしまうことは適切でない。そこで各EDUに対し、EDUに含まれる各単語ノードと接続されるようなMini-Hubノードを導入し、これがEDUノードと対応するようにした。以上により生成される文章グラフは、図1に示したように大きく二つの階層を持つ。これにより文章を持つ階層構造も適切に表現できると考えられる。

3.2 読解モデル: Graph_RC

本研究で提案する読解問題を解くモデルGraph_RCの概要を図5に示す。多肢選択型読解問題における文章、質問文、回答の選択肢1つを与

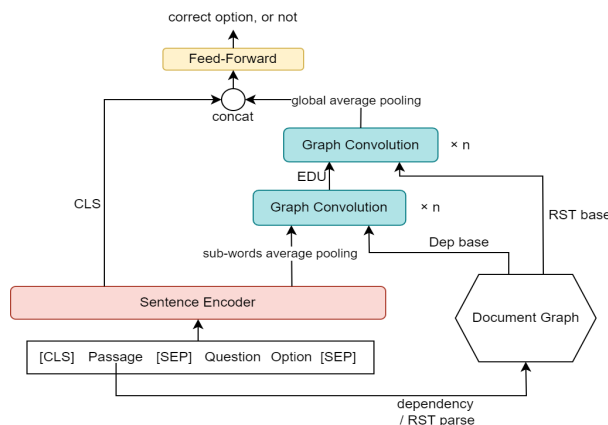


図5 読解モデル (Graph_RC)

え、正しい選択肢か否かの評価値を予測し、最終的に4つの選択肢の評価値にsoftmaxを施すことで正しい選択肢を求める。

Sentence Encoder(言語モデル)への入力は、[CLS]+文章+[SEP]+質問+回答の選択肢+[SEP]という形式で与える。また、3.1で述べた手順で文章のグラフ表現を生成しておく。言語モデルの出力として得られる、文脈を考慮したトークンごとの表現を文章グラフの下位階層の単語ノードに初期潜在表現として与える。ここで、文章グラフの単語ノードに対応するトークンが複数のサブワードトークンに分割されている場合は、これらの複数のトークンの表現を平均した結果を用いる。

グラフ畳み込みモデルによる文章グラフのエンコードにおいては、大きく二つの階層をもつグラフの構造(図1)に合わせ、各文の構造を表す下位階層の部分をエンコードした後、文章構造を表す上位階層の部分をエンコードすることで文章全体の表現を得る。上位階層のエンコードにおいては、下位階層のエンコードの結果Mini-Hubノードに集約された表現をEDUノードの初期表現として与え、畳み込み演算後の全EDUノードに対しglobal attention poolingを適用することで文章グラフの表現を得る。こうした階層的な処理により、GCNの畳み込み演算後に多くのノード全てを一度にpoolingすることでノードの表現が平均化され情報が失われる影響を低減する効果が期待できる。

最終的に言語モデルから得られるCLSトークンの表現と、グラフ畳み込みモデルから得られる文章グラフの表現を結合し、全結合層からなる出力層に与えることにより、注目している選択肢が回答であるかを判定する。

4 実験

4.1 実験設定

多肢選択型の機械読解データセットの中でも比較的データ数が豊富な RACE [7] を用いて評価実験を行った。RACE においては一つの文章に対して複数の質問と各質問ごとに4つの選択肢が与えられており、各質問に対して正しい選択肢を選ぶことが求められる。

実験においては Sentence Encoder として利用する言語モデルに BERT [1] を採用し、bert-base-uncased³⁾ と bert-large-cased-whole-word-masking⁴⁾ の二種類の事前学習済モデルを fine-tuning する形で利用した。評価指標には正答率を用いた。

図 5 におけるグラフ畳み込みモデルとしては RGCN[16] を用い、下位階層と上位階層は、それぞれ6層、4層とした。また、出力の全結合層は2層とした。BERTの最大系列長は512で統一し、[1]を参考にオプティマイザとして重み減衰に L2 正則化を適用した adam である adamw を使用した。

4.2 結果と考察

表 1 に各モデルの正答率を示す。ここで、Turkers とはランダムにサンプリングされたテストデータに対する人による正答率の平均であり、平均的な人の能力を表す。一方、Ceiling とは曖昧な問題や正解のない問題を除いた人間が回答可能と考えられる問題の割合であり、人間による正答率の上限値を近似する。Graph_RC-base、Graph_RC-large と記したのが本研究の提案モデルであり、それぞれ言語モデルに BERT-base、BERT-large を採用した結果である。また、表の CSFN、SG-Net というモデルは、文章の構造情報を言語モデルに併用する形で用いる既存研究であり、それぞれ言語モデルとして BERT-base、BERT-large を採用している。

表 1 から、BERT 単体を用いた場合と比べて、提案モデル Graph_RC が高い精度を達成したことが分かる。これにより、文章の構造情報を利用する有効性が確認できた。また意味役割の情報を利用する CSFN や単語依存構造を利用する SG-Net といった既存手法と比較しても精度向上が確認できたことから、文内に限らない異なる階層の構造情報を複合的

表 1 実験結果

モデル	正答率 [%]
Turkers[7]	73.3
Ceiling[7]	94.5
BERT-base[5]	65.0
CSFN[5]	67.7
Graph_RC-base	68.0
BERT-large[4]	72.0
SG-Net[4]	74.2
Graph_RC-large	75.1
random[7]	24.6

に用いることが有効だと考えられる。

Graph_RC のさらなる精度向上のためには、現状グラフ表現と言語モデルからの文脈表現の結合において単純な concat を適用している部分に、[5] で用いられているようなゲート機構やアテンションを導入することが考えられる。また現段階では、質問文や答えの選択肢については文章グラフに含まれていないため、これらの情報を適切に文章グラフに組み込むことも有効であると期待できる。

なお、入力する文章グラフや利用するグラフ畳み込みモデルの影響について調べるため、Graph_RC-base を用いて、RACE データセットの一部である RACE-M にて条件を変化させて学習・評価した場合についての分析を Appendix に掲載している。

5 おわりに

本研究では、機械読解タスクにおける精度向上を目的とし、グラフ構造による文章表現を利用する機械読解手法を提案した。具体的には、単語依存構造をベースに、文の位置関係や共参照、談話修辭構造の解析結果を用いることにより、文レベル、文章レベルの2つの階層からなる文章のグラフ表現を構成した。評価実験の結果、言語モデル単体を用いた場合と比べて精度向上が達成され、多肢選択型機械読解タスクにおいて、提案する文章グラフを用いて文章の構造情報を明示的に与えることの効果が確認できた。また既存の文章の構造情報を部分的に利用する手法との比較や、入力するグラフを変化させての分析の結果から、複数の異なる階層の構造情報を複合的に用いることの有用性が示唆された。

今後の課題としては、[17] など別の読解データセットや [18] などより大規模な言語モデルを用いた実験が考えられる。

3) <https://huggingface.co/bert-base-uncased>

4) <https://huggingface.co/bert-large-cased-whole-word-masking>

参考文献

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. **arXiv preprint arXiv:1810.04805**, 2018.
- [2] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. **Advances in neural information processing systems**, Vol. 32, , 2019.
- [3] Thang M Pham, Trung Bui, Long Mai, and Anh Nguyen. Out of order: How important is the sequential order of words in a sentence in natural language understanding tasks? **arXiv preprint arXiv:2012.15180**, 2020.
- [4] Zhuosheng Zhang, Yuwei Wu, Junru Zhou, Sufeng Duan, Hai Zhao, and Rui Wang. Sg-net: Syntax-guided machine reading comprehension. In **Proceedings of the AAAI Conference on Artificial Intelligence**, Vol. 34, pp. 9636–9643, 2020.
- [5] Qianwei Duan, Jun Huang, and Huiyan Wu. Contextual and semantic fusion network for multiple-choice reading comprehension. **IEEE Access**, Vol. 9, pp. 51669–51678, 2021.
- [6] 小坂直輝, 小林哲則, 林良彦. グラフによるテキスト表現の機械読解における適用性評価. 人工知能学会全国大会論文集, Vol. JSAI2021, pp. 3J1GS6a05–3J1GS6a05, 2021.
- [7] Guokun Lai, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy. Race: Large-scale reading comprehension dataset from examinations. **arXiv preprint arXiv:1704.04683**, 2017.
- [8] Shuyi Wang, Hui Song, Bo Xu, and Hongkuan Zhang. Enhancing embedding via two-level features for machine reading comprehension. In **China Conference on Knowledge Graph and Semantic Computing**, pp. 261–273. Springer, 2020.
- [9] Yawei Sun, Gong Cheng, and Yuzhong Qu. Reading comprehension with graph-based temporal-casual reasoning. In **Proceedings of the 27th International Conference on Computational Linguistics**, pp. 806–817, 2018.
- [10] Mrinmaya Sachan and Eric Xing. Machine comprehension using rich semantic representations. In **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**, pp. 486–492, 2016.
- [11] Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. Abstract meaning representation for sembanking. In **Proceedings of the 7th linguistic annotation workshop and interoperability with discourse**, pp. 178–186, 2013.
- [12] Sunil Kumar Sahu, Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. Inter-sentence relation extraction with document-level graph convolutional neural network. **arXiv preprint arXiv:1906.04684**, 2019.
- [13] Kevin Clark and Christopher D Manning. Deep reinforcement learning for mention-ranking coreference models. **arXiv preprint arXiv:1609.08667**, 2016.
- [14] William C Mann and Sandra A Thompson. **Rhetorical structure theory: A theory of text organization**. University of Southern California, Information Sciences Institute Los Angeles, 1987.
- [15] Vanessa Wei Feng and Graeme Hirst. A linear-time bottom-up discourse parser with constraints and post-editing. In **Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 511–521, 2014.
- [16] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In **European semantic web conference**, pp. 593–607. Springer, 2018.
- [17] Daniel Khashabi, Snigdha Chaturvedi, Michael Roth, Shyam Upadhyay, and Dan Roth. Looking beyond the surface: A challenge set for reading comprehension over multiple sentences. In **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)**, pp. 252–262, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.
- [18] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. **arXiv preprint arXiv:1907.11692**, 2019.
- [19] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. **arXiv preprint arXiv:1710.10903**, 2017.

A Appendix

本節には、Graph_RC-base を用いて、RACE データセットの一部である RACE-M にて条件を変化させて学習・評価した場合についての分析を記す。

A.0.1 文章グラフを変化させた場合の影響

表 2 入力グラフの比較 (3edge は隣接文, HUB ノード, 共参照による文グラフの統合を表す)

入力グラフ	正答率 [%]
Dep, 3edge, RST	68.4
Dep, 3edge	67.7
Dep	67.3

入力する文章グラフを変化させた場合の比較を表 2 に示す。表の上から、提案する文章グラフのすべての要素を用いた場合、下位階層の部分のみを用いた場合、下位階層のさらに単語依存構造グラフのみを用いた場合である。表 2 から提案する文章グラフ全体を用いた場合の精度が最も高く複数の構造情報を複合的に用いることが有用であると考えられる。また単語依存構造グラフのみを用いた場合と比較して、3 種類のエッジによる統合・補強を行い、文の位置情報や共参照情報を含んだグラフの方を用いた場合の方が精度が高いことも確認できたがその精度向上幅は僅かであった。これは文グラフの統合・補強用の 3 種類のエッジが文グラフに含まれる依存構造エッジと比較し少ないためだと考えられる。

表 3 は 1 つの文章グラフに含まれているエッジの大まかなタイプごとの平均本数を示したものであるが、文章に含まれる単語数に対応する依存構造エッジに比べて、文の数に対応する隣接文エッジや HUB 統合エッジ、共参照エッジは本数が少ないことがわかる。なお、共参照エッジに関しては今回用いた共参照解決エンジンが読解問題の文章のような複数文からなる長い文章に対して十分な精度を達成できておらず、全ての共参照を捉えきれなかったり、誤った共参照解決が行われている例も見受けられた。そのため異なる共参照解決器を用いて共参照解決自体の精度を向上させることで読解精度向上に寄与することが期待できる。

表 3 エッジの分布

エッジタイプ	Dep	next	hub	coref	RST
1 文章の平均エッジ数	322.8	18.8	19.8	18.6	38.5

A.0.2 GCN を変化させた場合の影響

利用するグラフ畳み込みモデルについて、RGCN[16] と GAT[19] を用いた場合の比較を表 4 に示す。表から RGCN を用いた場合の方が精度が高いことがわかり、文章グラフに存在するエッジの向きやラベルといった情報を直接利用できることが精度向上に寄与したと考えられる。

図 6 は文章グラフの上位階層のグラフに利用している談話修辭構造の修辭関係を表すエッジの種類ごとの出現頻度を表したものであり、全部で 19 種類のエッジが存在する。依存構造グラフの依存関係を表すエッジに関してはその種類はさらに多く計 45 種類のエッジが存在する。RGCN における畳み込み演算の基底数や GAT のアテンション計算におけるアテンションのマルチヘッドの数といったパラメータによっても変化すると考えられるが、こうした細かい関係の考慮は GAT より RGCN の方が優れていると推測される。一方で図 6 からは RST のエッジの分布には大きな偏りがあることがわかり、これは依存関係エッジでも同様である。RGCN を用いた場合、エッジの種類数が多すぎると出現頻度の低いエッジについて十分に学習が行われず精度低下をまねく可能性も考えられ、また計算コストの増大といった問題もある。そのため出現頻度の低いエッジについては同一のラベルにまとめるなど、文章グラフにおけるエッジのラベル付けの仕方についても検討の余地がある。

表 4 グラフ畳み込みモデルの比較

GCN layer	正答率 [%]
RGCN	68.4
GAT	66.0

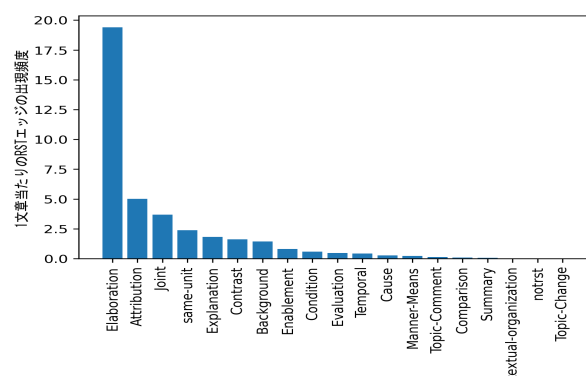


図 6 RST エッジの分布