

会話によるニュース記事伝達のための談話構造解析

高津 弘明¹安藤 涼太²松山 洋一¹小林 哲則¹¹ 早稲田大学² 内外切抜通信社

takatsu@pcl.cs.waseda.ac.jp

ando@naigaipc.co.jp

matsuyama@pcl.cs.waseda.ac.jp

koba@waseda.jp

1 はじめに

文間の依存構造を解析する手法および、結束性の高い文のまとまり(チャンク)を検出する手法と談話関係を分類する手法を提案する。

一貫性のある要約 [1, 2, 3] や対話 [4, 5] を実現するために、談話構造が制約として用いられる。我々は、ニュース記事の内容を効率的に伝える音声対話システム [6] においても、話を破綻させないために談話構造に従った情報伝達が重要であると考え、文を談話単位として、日本語のニュース記事に対して談話依存構造を付与したデータセットを構築した [7]。このデータセットは、ウェブニュースのクリッピングの専門家が 1200 個のニュース記事に対して「係り先」「談話関係」「チャンク」のアノテーションを行ったものである。係り先は、ルートノードから親ノード文までの情報が対象の文を理解するうえで必要最小限の情報になるように付与したものである。談話関係は、子ノード文が親ノード文に対してどのような意味の関係にあるかを分類したものである。チャンクは、親ノード文の内容を正しく理解するために欠かせない情報が子ノード文に書かれている場合、これらをまとめて提示すべきであることを表したものである。

本研究では、「係り先」「談話関係」「チャンク」を推定するモデルを提案し、作成したデータセットを用いて評価した結果を報告する。文間の係り受け解析については、BERT [8] による文の埋め込み表現を双方向の GRU [9] に与え、head selection モデル [10] により係り先を推定するモデルを提案する。談話関係分類とチャンク検出については、談話依存構造木をルートノードから葉ノードまでの文系列に分解し、各タスクの系列ラベリング問題をマルチタスクで学習するモデルを提案する。

本稿の構成は次の通りである。2 章で関連研究に

ついて述べる。3 章で文の係り先を推定するモデルについて説明する。4 章でチャンクと談話関係を推定するモデルについて説明する。5 章でデータセット [7] を用いてモデルの性能を評価した結果を報告する。6 章でまとめと今後の展望について述べる。

2 関連研究

談話構造解析は、文章を構成する文や節の間に成り立つ関係を解析する自然言語処理の基本的なタスクである。談話構造解析の結果は、文書要約 [3] や質問応答 [11]、機械翻訳 [12]、評判分析 [13] などの下流タスクのアプリケーションで用いられる [14]。

談話構造解析のための代表的なデータセットとして、RST Discourse Treebank [15] があり、文章から修辭構造木を自動で構築する手法が提案されてきた [16, 17, 18]。一方で、文書要約 [1, 2, 19, 20] のようなタスクへの応用を考えた場合、修辭構造木のような句構造よりも談話単位間の親子関係を直接表せる依存構造の方が望ましいと考えられ、修辭構造木を談話依存構造木へ変換する手法が提案された [19, 21]。また、変換アルゴリズムによって生成される談話依存構造木の性質が異なることから [22]、Yang らは、科学論文のアブストラクトに対して人手で EDU 間の依存構造と談話関係をアノテーションする方法を提案し、データセットとして SciDTB を構築した [23]。また、Yang らは、EDU 間の依存構造を解析する手法として、transition-based の手法 [24] と graph-based の手法 [21] を比較し、SciDTB を用いた場合、transition-based の手法の方が高い性能を示すことを確認した。単語間の係り受け解析においては、head selection モデルの有効性が確認されており [10]、本研究では、これを文間の係り受け解析モデルへ拡張する。

近年、自然言語処理の分野では、ラベルなしの膨大なテキストデータを使用して事前学習した言語モ

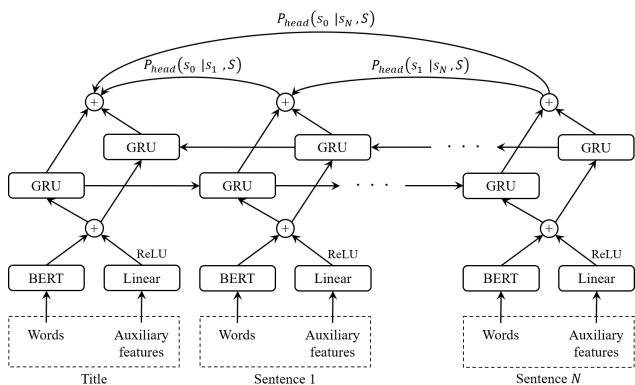


図1 係り先推定モデル

デルを下流タスクでファインチューニングするアプローチが注目されている [25]. 談話関係分類タスクでも, BERT [8] が暗黙的な談話関係の分類に有効であることが確認されており [26, 27], 本研究でも BERT 由来の文や単語の特徴量を利用する.

3 文間係り受け解析

提案モデルの概略図を図 1 に示す. 記事のタイトルを ROOT として加えた文系列 $S = (s_0 = \text{ROOT}, s_1, \dots, s_N)$ を入力とする. 文の単語系列を BERT [8] に与え [CLS] トークンに対応する最上位層の埋め込み表現を得る. これと文の補助情報 (文書内文位置, 文書内段落位置, 段落内文位置) の埋め込み表現を結合したベクトルを双方向の GRU [9] に与える. i 番目の文に対応する GRU の順方向と逆方向の隠れ層の出力を結合したベクトルを h_i とする. head selection モデル [10] に基づき s_j が s_i の係り先である確率 $P_{\text{head}}(s_j | s_i, S)$ を以下の式で計算する.

$$P_{\text{head}}(s_j | s_i, S) = \frac{\exp(g(h_j, h_i))}{\sum_{k=0}^N \exp(g(h_k, h_i))} \quad (1)$$

$$g(h_j, h_i) = v_h^T \tanh(U_h h_j + W_h h_i) \quad (2)$$

v_h, U_h, W_h は重みパラメータである.

4 談話関係分類とチャンク検出

談話依存構造木をルートノードから葉ノード (以下, root-to-leaf) までの文系列に分解し, この文系列に対する系列ラベリングにより, 談話関係とチャンクラベルを推定する. 提案モデルの概略図を図 2 に示す. マルチタスク学習の全体損失関数 L_{all} を談話関係分類タスクの損失関数 L_r とチャンク検出タスクの損失関数 L_c の重み付き和で定義する.

$$L_{\text{all}} = \lambda_r \times L_r + \lambda_c \times L_c \quad (3)$$

λ_r と λ_c は各タスクの重み係数である.

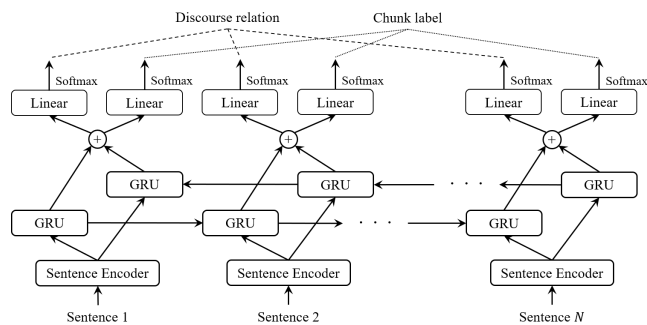


図2 談話関係・チャンク推定モデル

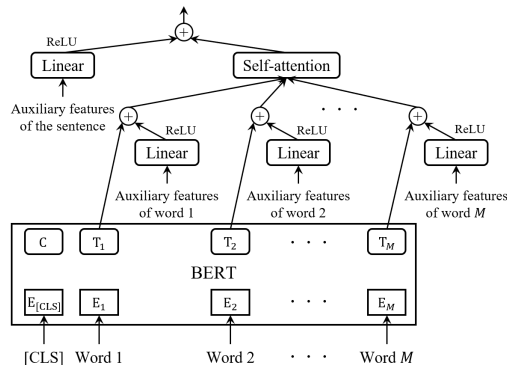


図3 Sentence Encoder

談話関係は, 子ノード文が親ノード文に対してどのような意味的關係にあるかを分類したものである. ラベルとして「開始」「結果」「原因」「背景」「呼応」「対比」「転換」「例示」「結論」「補足」があり, これらを softmax で識別する.

チャンクは, 親ノード文の内容を正しく理解するために欠かせない情報が子ノード文に書かれている場合, これらをまとめて提示すべきであることを表した文の集合である. チャンクのラベルとして, チャンクの開始を表す“B”, チャンクの内側を表す“T”, チャンクの終了を表す“E”, チャンクの外側を表す“O”を定め, これらを softmax で識別する.

談話関係を識別する手掛かりとして接続詞などの単語の情報が有効だと考えられる. そこで, Sentence Encoder では, BERT [8] の単語埋め込み表現と単語の補助情報の埋め込み表現を結合したものに対して self-attention [28] を計算する. さらに, 得られたベクトルと文の補助情報 (文書内文位置, 文書内段落位置, 段落内文位置, 談話依存構造木における深さ) の埋め込み表現を結合する. また, ネガティブな事象の原因はネガティブであることが多い [29] というように感情極性情報も談話関係の判断に有効だと考え, 単語の補助情報として品詞や活用形などの他, 単語の感情極性情報を与える.

表 1 文間係り受け解析性能：位置情報の効果

| | 位置情報なし | 位置情報あり |
|----------|--------|--------|
| Accuracy | 0.717 | 0.768 |

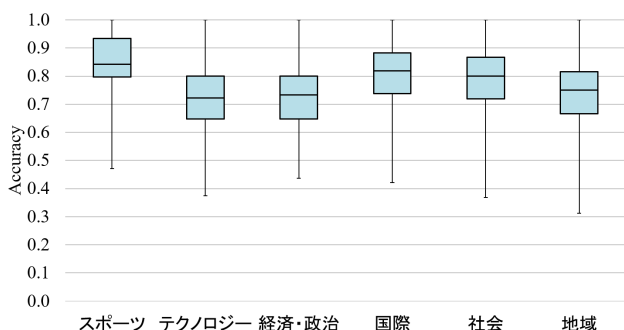


図 4 文間係り受け解析性能：ジャンルごとの性能

5 実験

ニュース記事に対して談話依存構造を付与したデータセット [7] を使用した。このデータセットは、15 文から 25 文のニュース記事 1200 個（6 ジャンル 200 記事ずつ）に対して、「係り先」「談話関係」「チャンク」を付与したデータセットである。

BERT の事前学習モデルとして、情報通信研究機構が公開しているモデルを用いた¹⁾。使用した BERT モデルは、日本語 Wikipedia の全記事に対して、Juman 辞書を用いた MeCab²⁾ [30] により形態素解析を行ったテキストを入力として、語彙数が 10 万語からなる BERT_{BASE} [8] を学習したモデルである。

単語の感情極性の辞書として、『日本語評価極性辞書』 [31, 32]³⁾、『単語感情極性対応表』 [33]⁴⁾、『意見（評価表現）抽出ツール用モデル』⁵⁾ の評価表現辞書および極性反転語の辞書を用いた。

評価は 10 分割交差検定で行い、ジャンルごとに記事数の 9/10 を訓練データ（1080 記事）、1/10 をテストデータ（120 記事）として分割した。

5.1 文間係り受け解析性能の評価

文の補助情報として文の位置情報を加えたときと加えなかったときの Accuracy を表 1 に示す。文の位置情報を加えたモデルは、位置情報を加えなかったモデルに比べて、Accuracy が 5% 以上高く、文の係り先の推定において、文の位置情報が有効であるこ

1) <https://alaginrc.nict.go.jp/nict-bert/index.html>
 2) <https://taku910.github.io/mecab/>
 3) <https://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resource%20Japanese%20Sentiment%20Polarity%20Dictionary>
 4) http://www.lr.pi.titech.ac.jp/~takamura/pndic_ja.html
 5) <https://alaginrc.nict.go.jp/li-outline.html#C-3>

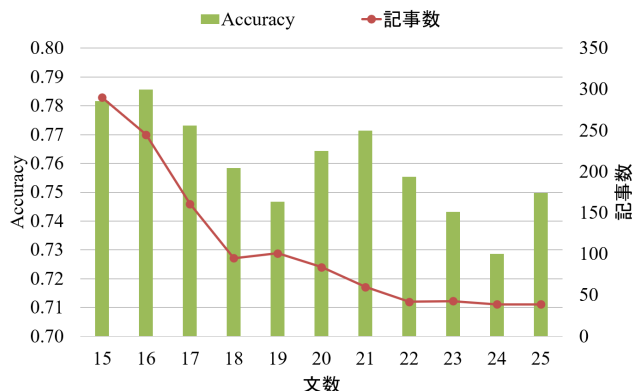


図 5 文間係り受け解析性能：記事の文数ごとの性能

とが分かった。

記事ごとに計算した Accuracy の分布をジャンルごとに箱ひげ図で表したものを図 4 に示す。ジャンルごとに比較するとスポーツ記事が最も高い性能を示した。スポーツ記事は試合の流れなどが線形につながった分岐の少ない木構造が多いため、係り先が推定しやすかったと考えられる。

記事の文数ごとの記事数および、各文数の記事ごとに計算した Accuracy の平均値を図 5 に示す。データ数の多い 15 文や 16 文の性能が他に比べて高い傾向を示したが、データ数の少ない 20 文以上の文数の記事においても Accuracy の減少は 5% 程度にとどまっていた。

5.2 談話関係分類性能の評価

データセットにおける談話関係の数に偏りがあるため、1200 記事のデータを全て使用すると低頻度の談話関係の識別性能が低くなる。そこで、データ数の偏りの影響を抑えるため、談話関係のそれぞれについて、対象の談話関係を 1 つ以上含む記事の root-to-leaf 文系列をデータセットとした。各談話関係を含む記事の数を以下に示す。「結果」:250, 「原因」:437, 「背景」:800, 「呼应」:578, 「対比」:459, 「転換」:205, 「例示」:396, 「結論」:685。その他、「開始」と「補足」があるが、「開始」は係り先がルートノードである文に自動的に付与され、「補足」は上記談話関係に分類されなかったものに対して付与されるため、これらは評価対象から外した。

各談話関係の分類性能 (F 値) を表 2 に示す。チャンク検出とのマルチタスク学習の結果は、 $(\lambda_r, \lambda_c) = (0.9, 0.1), (0.8, 0.2), \dots, (0.1, 0.9)$ のうち、最大のものを示している。シングルタスクモデルに比べてマルチタスクモデルの方が高い性能を示した。

表2 談話関係分類性能 (F 値)

| | シングルタスク | マルチタスク |
|----|---------|--|
| 結果 | 0.465 | 0.497 ($\lambda_r = 0.8, \lambda_c = 0.2$) |
| 原因 | 0.615 | 0.640 ($\lambda_r = 0.9, \lambda_c = 0.1$) |
| 背景 | 0.505 | 0.510 ($\lambda_r = 0.9, \lambda_c = 0.1$) |
| 呼応 | 0.406 | 0.417 ($\lambda_r = 0.9, \lambda_c = 0.1$) |
| 対比 | 0.888 | 0.896 ($\lambda_r = 0.5, \lambda_c = 0.5$) |
| 転換 | 0.678 | 0.696 ($\lambda_r = 0.6, \lambda_c = 0.4$) |
| 例示 | 0.410 | 0.466 ($\lambda_r = 0.8, \lambda_c = 0.2$) |
| 結論 | 0.442 | 0.449 ($\lambda_r = 0.9, \lambda_c = 0.1$) |

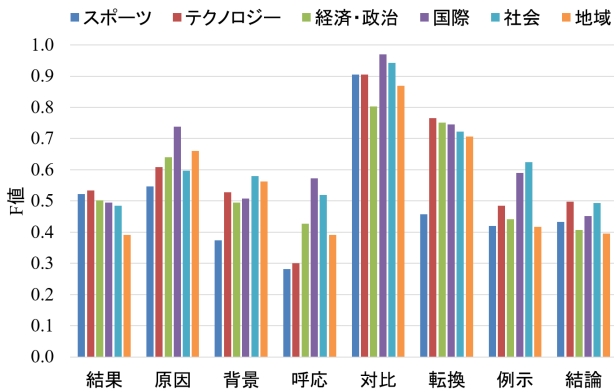


図6 談話関係分類性能: ジャンルごとの性能

談話関係ごとに比較すると、「対比」が最も高い性能を示した。「対比」が付与されている文は、「一方」や「ただ」のような接続詞で始まることが多く、これらが手掛かりになったと考えられる。一方、「例示」を表す接続表現として「例えば」や「具体的には」などが挙げられるが、実際に「例示」がアノテーションされた文は、このような接続表現を伴うものが少なく、「補足」との区別が付きにくかったため、「例示」の分類性能は低かったと考えられる。

ジャンルごとの各談話関係の分類性能を図6に示す。データセットにおいて、国際記事は他のジャンルに比べて「原因」が多く、スポーツ記事は「転換」が少ないというように、ジャンルごとの談話関係の出現頻度も性能に影響していると考えられる。

5.3 チャンク検出性能の評価

チャンクを少なくとも1つ含む595記事を使用した。データセット[7]には、チャンクとして、強いチャンクと弱いチャンクの2種類がアノテーションされているが、これらを区別せずに推定した。

チャンク検出性能とチャンク文検出性能の2つの観点で評価した。チャンク検出性能は、チャンクの範囲が全て合っているものを正解としたときのF値である。チャンク文検出性能は、B,EラベルをIラベルに集約し、チャンクに含まれる文か否かの2値

表3 チャンク検出性能 (F 値)

| | シングルタスク | マルチタスク |
|-----------|---------|--------|
| チャンク検出性能 | 0.605 | 0.629 |
| チャンク文検出性能 | 0.720 | 0.737 |

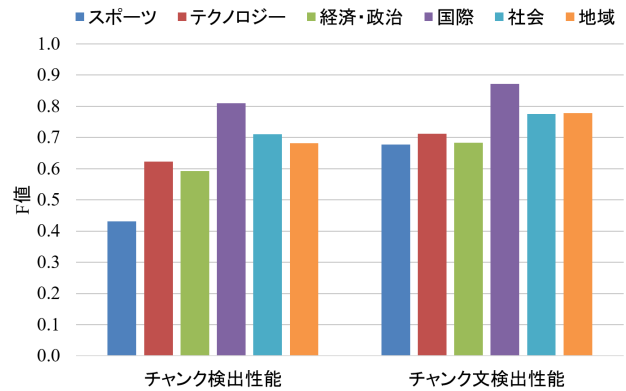


図7 チャンク検出性能: ジャンルごとの性能

ラベルに修正したときの、Iラベルに関するF値である。なお、チャンクは2文以上からなるため、Iラベルが単体で出現していた場合は、ルールでOラベルに修正した。

チャンクをシングルタスクで学習したときの結果と、チャンクと談話関係をマルチタスクで学習したときの結果 ($\lambda_r = 0.2, \lambda_c = 0.8$) を表3に示す。チャンクと談話関係をマルチタスクで学習したモデルの方が高い性能を示した。談話関係を一緒に学習することで、「対比」関係の文が弱いチャンクになりやすいなどの傾向を学習できたためだと考えられる。

ジャンルごとの性能を図7に示す。国際記事は、弱いチャンクに出現しやすい「一方」や「ただ」などの接続詞が多く出現することから、チャンク検出性能が他に比べて高かったと考えられる。スポーツ記事は、試合の流れなどの長いチャンクが多く、チャンクを部分的には当てられているが、チャンクの開始から終わりまでを当てるのが難しかったため、チャンク文検出性能に比べてチャンク検出性能が低かったと考えられる。

6 おわりに

文の「係り先」「談話関係」「チャンク」を推定するモデルを提案し、日本語のニュース記事に対してこれらをアノテーションしたデータセット[7]を用いて性能を評価した。

今後は、談話構造を制約とした要約[34]や対話のアルゴリズムについて検討する。

謝辞 本研究は、JST START (JPMJST1912) の支援を受けたものである。

参考文献

- [1] Yuta Kikuchi, Tsutomu Hirao, Hiroya Takamura, Manabu Okumura, and Masaaki Nagata. Single document summarization based on nested tree structure. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 315–320, 2014.
- [2] Tsutomu Hirao, Masaaki Nishino, Yasuhisa Yoshida, Jun Suzuki, Norihito Yasuda, and Masaaki Nagata. Summarizing a document by trimming the discourse tree. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, Vol. 23, No. 11, pp. 2081–2092, 2015.
- [3] Jiacheng Xu, Zhe Gan, Yu Cheng, and Jingjing Liu. Discourse-aware neural extractive text summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5021–5031, 2020.
- [4] Song Feng, Kshitij Fadnis, Q. Vera Liao, and Luis A. Lastras. DOC2DIAL: a framework for dialogue composition grounded in business documents. In *Proceedings of the 33rd Conference on Neural Information Processing Systems*, p. 1.
- [5] Mihai Rotaru and Diane J. Litman. Exploiting discourse structure for spoken dialogue performance analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 85–93, 2006.
- [6] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則. 意図性の異なる多様な情報行動を可能とする音声対話システム. *人工知能学会論文誌*, Vol. 33, No. 1, pp. 1–24, 2018.
- [7] 高津弘明, 安藤涼太, 本田裕, 松山洋一, 小林哲則. ニュース記事に対する談話構造と興味度のアノテーション～ニュース対話システムのパーソナライズに向けて～. *言語処理学会 第 27 回年次大会 発表論文集*, 2021.
- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171–4186, 2019.
- [9] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734, 2014.
- [10] Xingxing Zhang, Jianpeng Cheng, and Mirella Lapata. Dependency parsing as head selection. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 665–676, 2017.
- [11] Peter Jansen, Mihai Surdeanu, and Peter Clark. Discourse complements lexical semantics for non-factoid answer reranking. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 977–986, 2014.
- [12] Junxuan Chen, Xiang Li, Jiarui Zhang, Chulun Zhou, Jianwei Cui, Bin Wang, and Jinsong Su. Modeling discourse structure for document-level neural machine translation. In *Proceedings of the 1st Workshop on Automatic Simultaneous Translation*, pp. 30–36, 2020.
- [13] Baptiste Chardon, Farah Benamara, Yannick Mathieu, Vladimir Popescu, and Nicholas Asher. Measuring the effect of discourse structure on sentiment analysis. In *Proceedings of the 14th international conference on Computational Linguistics and Intelligent Text Processing*, pp. 25–37, 2013.
- [14] Shafiq Joty, Giuseppe Carenini, Raymond T. Ng, and Gabriel Murray. Discourse analysis and its applications. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pp. 12–17, 2019.
- [15] Lynn Carlson, Daniel Marcu, and Mary Ellen Okurovsky. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, pp. 1–10, 2001.
- [16] Longyin Zhang, Yuqing Xing, Fang Kong, Peifeng Li, and Guodong Zhou. A top-down neural architecture towards text-level parsing of discourse rhetorical structure. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 6386–6395, 2020.
- [17] Nan Yu, Meishan Zhang, and Guohong Fu. Transition-based neural RST parsing with implicit syntax features. In *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 559–570, 2018.
- [18] Qi Li, Tianshi Li, and Baobao Chang. Discourse parsing with attention-based hierarchical neural networks. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 362–371, 2016.
- [19] Tsutomu Hirao, Yasuhisa Yoshida, Masaaki Nishino, Norihito Yasuda, and Masaaki Nagata. Single-document summarization as a tree knapsack problem. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1515–1520, 2013.
- [20] Yasuhisa Yoshida, Jun Suzuki, Tsutomu Hirao, and Masaaki Nagata. Dependency-based discourse parser for single-document summarization. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1834–1839, 2014.
- [21] Sujian Li, Liang Wang, Ziqiang Cao, and Wenjie Li. Text-level discourse dependency parsing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 25–35, 2014.
- [22] Katsuhiko Hayashi, Tsutomu Hirao, and Masaaki Nagata. Empirical comparison of dependency conversions for RST discourse trees. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 128–136, 2016.
- [23] An Yang and Sujian Li. SciDTB: Discourse dependency treebank for scientific abstracts. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 444–449, 2018.
- [24] Yizhong Wang, Sujian Li, and Houfeng Wang. A two-stage parsing method for text-level discourse analysis. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 184–188, 2017.
- [25] XiPeng Qiu, TianXiang Sun, YiGe Xu, YunFan Shao, Ning Dai, and XuanJing Huang. Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, Vol. 63, No. 10, pp. 1872–1897, 2020.
- [26] Yudai Kishimoto, Yugo Murawaki, and Sadao Kurohashi. Adapting BERT to implicit discourse relation classification with a focus on discourse connectives. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 1152–1158, 2020.
- [27] Wei Shi and Vera Demberg. Next sentence prediction helps implicit discourse relation classification within and across domains. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 5790–5796, 2019.
- [28] Zhouhan Lin, Minwei Feng, Cícero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. A structured self-attentive sentence embedding. In *Proceedings of the 5th International Conference on Learning Representations*, pp. 1–15, 2017.
- [29] Jong-Hoon Oh, Kentaro Torisawa, Chikara Hashimoto, Takuya Kawada, Stijn De Saeger, Jun’ichi Kazama, Yiu Wang. Why question answering using sentiment analysis and word classes. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 368–378, 2012.
- [30] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to Japanese morphological analysis. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 230–237, 2004.
- [31] Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, Kenji Tateishi, and Toshikazu Fukushima. Collecting evaluative expressions for opinion extraction. In *Proceedings of the 1st International Joint Conference on Natural Language Processing*, pp. 596–605, 2004.
- [32] 東山昌彦, 乾健太郎, 松本裕治. 述語の選択選好性に着目した名詞評価極性の獲得. *言語処理学会 第 14 回年次大会 発表論文集*, pp. 584–587, 2008.
- [33] Hiroya Takamura, Takashi Inui, and Manabu Okumura. Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pp. 133–140, 2005.
- [34] 高津弘明, 安藤涼太, 本田裕, 松山洋一, 小林哲則. 談話構造制約付きパーソナライズド抽出型要約. *言語処理学会 第 27 回年次大会 発表論文集*, 2021.