根拠情報抽出の課題設計と予備実験*

飯田 龍 †　　乾 健太郎 ‡　　松本 祐治 ‡
† 東京工業大学 大学院情報理工学研究科
ryu-i@cl.cs.titech.ac.jp
‡ 奈良先端科学技術大学院大学 情報科学研究科
{inui,matsu}@is.naist.jp

1 はじめに

本稿では，文章に記述された意見や命題などの根拠を抽出する問題を談話構造解析の部分問題とみなし，その問題設計と自動解析の結果について報告する．この根拠抽出の技術が実現することで，例えば，意見情報抽出 [17] の課題で抽出された ⟨ 対象，属性，評価 ⟩ のような意見の構成素に対して，なぜそのような意見が記述されたかという根拠をユーザ（意見の読み手）に提示することができる．これにより，ユーザは意見（帰結）の妥当性・信憑性を根拠側の情報から判断することが可能となる．例えば，例 (1) では，(a)⟨ ○○，ピザ，焼き立て ⟩，(b)⟨ ○○，ピザ，おいしい ⟩，(c)⟨ ○○，パスタ，おいしくない ⟩，(d)⟨ ○○，ϕ，行かない（という書き手の判断）⟩ などの意見の断片が抽出の対象となるが，読み手がこの記事を読み，「ピザを注文すべきか否か」を判断する場合，もしくは「その店へ行くべきかどうか」の決断を行う際には，(a) や (c) に記述された内容がそれぞれ (b) や (d) に関する根拠として参照すべき事実や書き手の判断を表していると考えられる．

(1) 昨日は○○というレストランに行ってきました．ピザは焼き立てでおいしかったけど，パスタはあまりおいしくなかった．もう行かないと思う．

上で述べたように，根拠情報の抽出の技術は（主観的な）意見の抽出課題（例えば，MPQA Corpus[14] や NTCIR 2007 blog[8]）で重要な要素技術となると考えられる．これに加え，読み手が “マイナスイオンは健康に良い” といった命題について真偽を判断するに場合にも重要であり，例えば，命題間の対立・含意・根拠の関係を可視化することを目的とする言論マップ [19] のような応用処理においても必須な要素技術である．

本研究で対象とする根拠関係の抽出は，既存研究である談話構造の部分解析とみなすことができる．談話関係の定義については，Mann らの修辞構造理論 [3] や Hobbs の定義した談話関係 [2] などがあり，最近では Penn Discourse Tree Bank[6] で接続表現ごとにどの範囲とどの範囲が関係するかといった情報が付与されるなどさまざまな定義があり，それぞれの定義に基づいたタグ付きコーパスが構築されている．また，PDTB については近年機械学習に基づく関係同定の試みが報告されており [12, 13]，談話関係の自動同定についても研究者の関心を集めている．

本研究では，既存研究ですでに定義されているさまざまな談話関係のうち，根拠情報が必要となるいくつかの応用処理を想定した場合に，どのような関係を抽出すべきかを議論し，その関係が既存研究と比較してどのようなものであるかを説明する．また，定義した問題に対し機械学習に基づく抽出手法を適用し，表層的な手がかりなど単純に抽出可能な情報のみでどの程度の抽出精度が得られるかを調査する．自動抽出には Wellner ら [12] と同様に述語項構造解析もしくは意味役割付与などで用いられているような手法を適用するが，この際，関係を根拠側から同定するかもしくは帰結側から同定するか，関係を同定する際の抽出手順など，いくつかの選択肢が存在する．そこで，今回の評価実験ではいくつかの項目について比較を行い，最も精度が良かった手法について誤り分析を行った．

本稿では，まず 2 節で根拠情報と関連する既存の談話関係を概観し，3 節で根拠情報タグ付きコーパスの構築について説明する．次に 4 節でタグ付与した情報の自動抽出とその誤り分析の結果を報告する．最後に 5 節でまとめる．

2 関連研究

今回対象とする文章中の談話片間の根拠-帰結関係を抽出する問題は，文章中の談話構造を部分的に解析する問題に対応すると考えることができる．近年の談話関係に関する研究では，修辞構造理論 (RST)[3] や Hobbs[2] の談話関係，Penn Discourse Tree Bank (PDTB) など種類は異なるが談話関係が付与されたタグ付きコーパスが構築され [1, 15, 6]，実際に自動解析を行った結果についても報告されている [4, 10, 12, 13]．この節では，今回対象としたい根拠-帰結関係が既存の談話関係の定義においてどのように扱われているかを概観する．

最初に RST については，根拠（evidence）関係が 1 節の例に特に関連しており，この関係を付与する際には表 1

*The Task Definition of Evidence-Conclusion Relation Extraction and its Preliminary Empirical Evaluation
Ryu Iida†, Kentaro Inui‡, and Yuji Matsumoto‡
† Tokyo Insititude of Technology
‡ Nara Institute of Science and Technology

表 1: RST における根拠関係の定義

nuclear(帰結) 側の制約: 書き手が満足できるほど読み手が帰結について信じていないかもしれない.
satellite(根拠) 側の制約: 読み手は根拠を信じられる，もしくは根拠に信憑性がある.
帰結と根拠の組み合わせについての制約: 読み手の根拠への理解が帰結についての信念を増す.
書き手の意図: 読み手の帰結への信念を増す.

にまとめる判断の基準を満している必要がある．表 1 からわかるように，RST においては根拠箇所で提示された情報が読み手が帰結箇所を信じるための情報の増加に影響しているかどうかが重要視されている．

一方，PDTB[6] では，主に明示的に出現している接続表現に対し，どのセグメントがどのセグメントと関係しているかを接続表現の項としてタグ付与している．例えば，例 (2) では接続詞 *After* に関して，“spending...September” を ARG1, “adjusting...inflation” を ARG2 として関係付ける．

(2) *After* [$_{arg2}$ adjusting for inflation] the Commerce Department said [$_{arg1}$ spending didn't change in September]

PDTB については，“because” などの接続表現について付与された関係が根拠情報に関連するが，PDTB の作業仕様では “:” や “;” などで区切られない同一文内の節の間や隣接しない文間などには関係が付与されないという網羅性の問題が残る [11]．また，PDTB の自動解析については，Wellner ら [12] が接続表現と対応する 2 つのセグメントを同定する問題を項同定の問題とみなし，既存の項同定・意味役割付与の手法を適用した結果について報告している．彼らの手法では，明示的に各談話セグメントの範囲を同定するのではなく，代わりに各セグメントの係り受け構造上の主辞を同定することで，接続表現の項となるセグメントの範囲を近似的に見積る．例えば，例 (2) の例については ARG1 と ARG2 についてそれぞれの主辞となる “adjusting” と “change” を同定する問題を解く．

3 根拠情報のタグ付与

3.1 根拠-帰結関係の仕様についての検討

本研究で対象とする根拠-帰結の関係は，意見情報抽出の技術で抽出された意見の根拠や，言論マップ [19] などの応用処理で必要となる命題の真偽を判断するための根拠など，既存の談話関係で言えば，原因・理由・根拠・動機・目的に相当する関係であり，これらは帰結側の表現がどのような表現であるかによって異なることがわかる．例えば，帰結側が “iPod touch には満足している” という意見の場合は “iPod touch が多機能である” ことがその意見を判断する際の根拠となる．一方，帰結側が “ 内閣支持率が低下している” という命題については “全国電話世論調査では，不支持が前回調査よりも 9 ポイント増加している” したという定量的な数値や “内閣の経済政策に対する不満が高まっている” という具体的な不満の内容などが根拠となる．また，帰結が “太郎が医者になった” という（過去に起こった）行為については，“ 太郎が数年前に医大に入学した” という行為や “ 病気で苦しむ人を救いたい” という動機を表す表現がその行為の根拠となる．

また，既存の談話構造の研究において定義された談話関係の観点から見ると，根拠の関係も RST が扱うような読み手が帰結側の内容についての信念が増すような場合にのみ関係を付与するという信念の度合いを吟味する狭義のものから，Marcu ら [5] がいくつかの既存研究に基づき分類した粒度の粗い Cause-Explanation-Evidence（CEV）関係などさまざまな関係が根拠情報とみなせることがわかる．

このような定義がある中で，我々は (a) 主観的意見についての根拠だけでなく客観的な行為・事象などについての原因・理由なども応用処理によっては需要があり，(b) 個別のアプリケーションについては帰結側を評価表現辞書[1]や主観・客観の判別，行為か否かの判別などによってほしい根拠の情報を取捨選択できる，という 2 つの理由より，抽出対象とする根拠-帰結関係を上述の例で示した行為・命題，判断についての原因・理由・動機・目的の関係として定義した．このため，“ため”，“ので”，“から” のような接続表現がこれらの根拠か否かを判断するための手がかりとなり，特に接続表現 “ため”，“ので” で関連付けられる 2 つの述語には必ず抽出対象となる[2]．

また，談話関係のタグ付与では関係を付与する範囲自体が問題となるが，今回の作業では Wellner ら [12] のような各セグメントの主辞のみを抽出対象とすることを想定し，根拠とその対応する帰結箇所の主辞を含む文節に人手でタグ付与を行う．主辞に該当する箇所は動詞，形容詞，名詞+“だ” などの述語が相当するが，“（⟨ 述語 ⟩ と）思う”，“（⟨ 述語 ⟩ ことを）計画する” などは広義のモダリティとみなし，思う（計画する）対象となる ⟨ 述語 ⟩ 側を主辞相当の表現とする．このような広義のモダリティを含め，述語該当箇所がどのようなテンス・アスペクト・モダリティで出現していても抽出対象とする．この結果に対し，例えば，原ら [18] の事実性解析を別途適用することで実際に起こった行為の根拠を抽出するなど，さまざまな用途に利用可能であると考えられる．

今回の作業内容をまとめると以下のようになる．

- **根拠帰結の関係は広く原因・理由・動機・根拠・目的などを含めた関係に付与**.
- **談話のセグメントは明示的に決定せず，主辞にタグ付与**.

3.2 作業方法と作業経過

5 億文コーパス [16] を文 ID に従って文章に復元し，その文章への根拠-帰結関係の人手タグ付与を行う．ただし，文章全体を対象にすると，遠距離で出現している根拠と帰結の対を考慮する必要があり，作業が困難にな

[1] http://www.syncha.org/evaluative_expressions.html

[2] ただし，述語とその必須項の間には根拠の関係を認めないよう定義した．例えば，“テロが飛行機の到着を遅らせた” で “テロ” が “飛行機が遅れた” ことの根拠とはみなさない．

る．一般に根拠該当箇所は帰結の近傍に出現するため，タグ付与の際にはあらかじめ“捕鯨問題”や“再販問題”，“iPod”など 9 種類のキーワードを用意しておき，そのキーワードを含む前後 2 文のみをタグ付与対象とした．文書集合全体のうち，作業者 1 人が 2954 の抜粋に対しタグ付与作業を行い，帰結 4333 箇所に対し根拠 4350 箇所にタグ付与された結果を得た．

4 自動抽出の評価

3 節で設計した根拠-帰結の関係をどの程度自動的に抽出できるかを調査するために評価実験を行った．付与された根拠-帰結関係のうち同一文内に出現している関係（全体の 94%，4053 事例）のみを対象に自動同定の評価を行う．評価は与えられた文中の任意の文節の組み合わせに対して根拠-帰結の関係となるかを同定する問題とする．また，“ため”，“から”，“ので”の 3 つの接続表現で係り受け関係となる 2 つの文節をすべて根拠-帰結の関係として抽出するモデルをベースラインモデルとし，機械学習に基づく手法と比較する．

4.1 根拠抽出手法

自動抽出の手法には Wellner ら [12] と同様に機械学習に基づく手法を採用する．彼らの談話関係同定の枠組みは，機械学習に基づく共参照解析の観点から見ると，Soon ら [9] の 2 値分類を用いた共参照解析の手法に概ね対応し，根拠（もしくは帰結）該当箇所を照応詞とみなし，帰結（根拠）該当箇所を先行詞とすることに相当する．共参照解析の問題では，先行詞候補集合の中から最初に最も先行詞らしい候補を決定し，その候補と照応詞の対を用いて照応関係になるか否かの分類問題を解くという 2 段階の処理を行うことで，Soon らの手法より高い精度で解析できる [21]．そこで，この 2 段階の処理を根拠抽出の処理に適用することで抽出精度に影響があるかについても調査を行う．帰結（根拠）側が与えられたときに対となる最も根拠（帰結）らしい候補の選別には，我々が以前提案したトーナメントモデル [20] を利用した．この 2 段階の手法ならびに Soon らの手法の詳細は文献 [21] を参照されたい．

素性抽出で必要となる形態素・係り受け解析には茶筌[3]，CaboCha[4]を利用した．また，分類器は SVMlight[5]を利用し，線形カーネル，パラメタ c は 1.0 に設定して評価を行った．

4.2 素性

学習・分類に用いる素性は以下に示す (a) 帰結（根拠）候補単体から抽出可能な素性と (b) 根拠-帰結の候補対から抽出できる素性の 2 種を用いた．

(a) **帰結（根拠）候補単体から抽出される素性**: 文頭か否か, 文末か否か, 主辞の品詞，主辞の見出し語，候補が述語か否か，候補文節内の機能語．

[3]http://chasen-legacy.sourceforge.jp/
[4]http://chasen.org/~taku/software/cabocha/
[5]http://www.cs.cornell.edu/people/tj/svm_light/

表 2: 帰結 (根拠) が存在する場合の根拠 (帰結) 同定の精度

モデル	精度
BM	0.540 (2188/4053)
1step:Ev→Cn	0.722 (2927/4053)
1step:Cn→Ev	0.808 (3276/4053)
2step:Ev→Cn	0.680 (2755/4053)
2step:Cn→Ev	0.814 (3298/4053)

BM:ベースラインモデル，1step:根拠と帰結を同時に決定するモデル（共参照解析における Soon ら [9] のモデルに相当），2step:帰結（根拠）側の文節が与えられら場合に根拠（帰結）を同定するモデル，Ev→Cn: 根拠文節から帰結側を同定，Cn→Ev: 帰結文節から根拠側を同定．

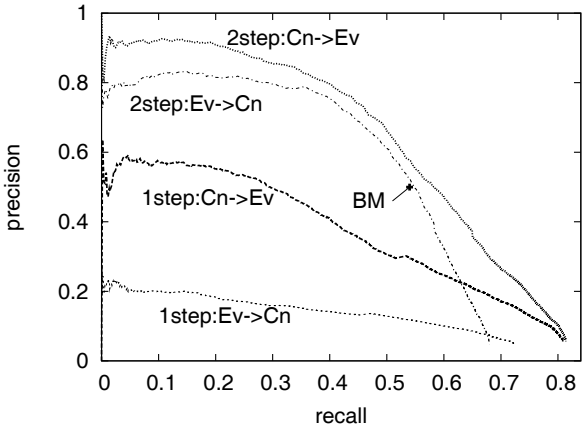

図 1: 各モデルの根拠-帰結関係の抽出結果

(b) **帰結候補と根拠候補から抽出される素性**: 根拠（帰結）候補が帰結（根拠）候補に係るか否か, 根拠（帰結）が帰結（根拠）より前に出現するか否か, 帰結候補と根拠候補の間の係り受けのパス中の主辞の品詞とすべての機能語．

4.3 実験結果

まず，帰結（もしくは根拠）の主辞を含む文節がどれであるかを与え，必ず対応する根拠（帰結）が存在する場合に，どのくらい正しく対応する根拠（帰結）文節を同定できるかを調査した．5 分割交差検定で評価した結果を表 2 に示す．この結果より，根拠該当箇所を与えられた場合に帰結側の曖昧性を解消するよりも，帰結から根拠を同定するほうの精度が良いことがわかる．これは，根拠側の文節には“ので”や“から”のような手がかり語を含むため，帰結側を同定するのに比べて問題が容易であったと考えられる．

次に，分類の際に出力されるスコア（分離平面からの距離）を信頼度とし再現率-精度曲線を描いた結果を図 1 に示す．この結果より，2 段階で帰結側から根拠箇所を同定する手法が最も高い精度であり，例えば再現率約 5 割で精度約 7 割の品質を得ていることがわかる．

4.4 誤り分析

解析を誤った事例をいくつかの観点で分析を行った．まず，表 2 のベースラインモデル，つまり接続表現のみを利用した規則ベースのモデルの再現率は約 56%であり，残りの 44%が抽出できていない．ベースラインで抽

出できなかった事例を人手で分析した結果，以下のような問題に起因することがわかった．

- 誤りの約 4 割は “ため”，“ので” のような接続表現を伴っていたにも関わらず，係り受け解析が誤っていたために対とすべき帰結側の同定を誤っていた．
- 上述の接続表現を伴う場合を除いた誤り事例（全体の約 6 割）のうち，根拠と帰結の文節が係り受け関係にあるのは誤り全体の約 35%であった．これらの事例のうちほとんどが “後継者が いなくなっていて根拠 困っているそうです帰結” や “ 実態が 明らかにされ,根拠 失望している帰結” のような連用中止やテ形接続で根拠と帰結が関係している事例であった．
- 残りの約 25%には上で述べた問題に加え，“死刑制度を 容認する帰結 する理由は... のため” のような倒置や，“ドラフト制度の 占める 意義が 大きいという根拠 ことを 理由に... と 考えた帰結” のような理由を表す表現の多様性など種々雑多な問題を含む．

また，過剰にモデルが関係を同定したため，図 1 に示すようにベースラインの精度は約 5 割という結果となっている．この原因としては，上述の係り受け誤りの影響による誤った根拠-帰結関係の同定が主なものであった．これらの誤り，特に係り受けに関する誤りについては，提案する機械学習に基づくモデルでは 4.2 で示した素性を用いることで，ある程度頑健に解析できていることがわかる．例えば，表 2 の結果より，対応する根拠がある場合は約 8 割の精度で候補群から根拠を同定可能なことがわかる．しかし，現状では根拠と帰結の関係となるか否かの分類の際に，自動獲得された述語間の因果・含意関係に関する知識などを利用していないため，提案手法の場合でも連用中止，テ形接続の事例については学習・分類のための情報が明らかに不足している．人手で整備された，もしくは自動獲得された述語間の関係知識などを導入することでどの程度この問題が改善されるかは次に取り組むべき興味深い課題だと考えられる．

5 おわりに

本稿では，日本語の根拠-帰結関係の抽出問題を設計し，実際にその関係を自動的に同定する試みについて報告した．5 億文コーパス [16] を用いて根拠タグ付きコーパスを作成し，このコーパスをもとに関係同定の予備実験を行い，単純な素性でどの程度の解析精度が得られるかを調査した．この結果，ある程度の同定精度を得られたが，さらなる品質向上のためには， 4.4 の誤り分析で述べたように，述語間の関係知識などの資源の適用を考えていく必要がある．

これらの課題に加え，実際の応用処理で根拠抽出の技術を利用するために，現在利用可能な資源で解析できる事例を機械的に選別するということも考えられる．例えば，one-class SVM[7] などを用い，分布の高密度領域を推定することで，解析対象をあらかじめ取捨選択するといった試みが考えられる．このような解析対象の選別に関する課題は今回扱った問題だけでなく，さまざまな応用処理に関連する要素技術において重要だと考えられ，この問題についても今後さらに検討していく予定である．

参考文献

[1] Carlson, L., Marcu, D. and Okurowski, M. E.: Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory, *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*, pp. 1–10 (2001).

[2] Hobbs, J. R.: On the coherence and structure of discourse, Technical Report 85-37, CSLI (1985).

[3] Mann, W. C. and Thompson, S. A.: Retorical Structure Theory: Toward a functional theory of text organization, *Text*, Vol. 8, No. 3, pp. 243–281 (1988).

[4] Marcu, D. and Echihabi, A.: An unsupervised approach to recognizing discourse relations, *Proceedings of ACL*, pp. 368–375 (2001).

[5] Marcu, D. and Echihabi, A.: An Unsupervised Approach to Recognizing Discourse Relations, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 368–375 (2002).

[6] Miltsakaki, E., Prasad, R., Joshi, A. and Webber, B.: The Penn Discourse Treebank, *Proccedings of the Language Resources and Evaluation Conference*, pp. 2237–2240 (2004).

[7] Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J. and Williamson, R. C.: Estimating the support of a high-dimensional distribution, TR 99-87, Microsoft Research (1999).

[8] Seki, Y., Evans, D. K., Ku, L., Sun, L., Chen, H. and Kando, N.: Overview of Multilingual Opinion Analysis Task at NTCIR-7, *Proceedings of NTCIR-7 Workshop Meeting*, pp. 185–203 (2008).

[9] Soon, W. M., Ng, H. T. and Lim, D. C. Y.: A Machine Learning Approach to Coreference Resolution of Noun Phrases, *Computational Linguistics*, Vol. 27, No. 4, pp. 521–544 (2001).

[10] Soricut, R. and Marcu, D.: Sentence Level Discourse Parsing using Syntactic and Lexical Information, *Proceedings of HLT-NAACL*, pp. 149–156 (2003).

[11] The PDTB Research Group: The Penn Discourse Treebank 2.0 Annotation Manual (2007).

[12] Wellner, B. and Pustejovsky, J.: Automatically Identifying the Arguments of Discourse Connectives, *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 92–101 (2007).

[13] Welwell, R. and Baldridge, J.: Discourse Connective Argument Identification with Connective Specific Rankers, *Processing of the IEEE International Conference on Semantic Computing*, pp. 198–205 (2008).

[14] Wiebe, J., Wilson, T. and Cardie, C.: Annotating expressions of opinions and emotions in language, *Language Resources and Evaluation*, Vol. 39, No. 2-3, pp. 165–210 (2005).

[15] Wolf, F. and Gibson, E.: Representing discourse coherence: A corpus-based analysis, *Computational Linguistics*, Vol. 31, No. 2, pp. 249–287 (2005).

[16] 河原大輔, 黒橋禎夫: 高性能計算環境を用いた Web からの大規模格フレーム構築, 情報処理学会 自然言語処理研究会報告 NL-171, pp. 67–73 (2006).

[17] 乾孝司, 奥村学: テキストを対象とした評価情報の分析に関する研究動向, 自然言語処理, Vol. 13, No. 3, pp. 201–241 (2006).

[18] 原一夫, 乾健太郎: 事態抽出のための事実性解析, 情報処理学会自然言語処理研究会報告 NL-183, pp. 75–80 (2008).

[19] 村上浩司, 松吉俊, 隅田飛鳥, 森田啓, 佐尾ちとせ, 増田祥子, 松本裕治, 乾健太郎: 言論マップ生成課題：言説間の類似・対立の構造を捉えるために, 情報処理学会自然言語処理研究会報告 NL-186, pp. 55–60 (2008).

[20] 飯田龍, 乾健太郎, 松本裕治: 文脈的手がかりを考慮した機械学習による日本語ゼロ代名詞の先行詞同定, 情報処理学会論文誌, Vol. 45, No. 3, pp. 906–918 (2004).

[21] 飯田龍, 乾健太郎, 松本裕治, 関根聡: 最尤先行詞候補を用いた日本語名詞句同一指示解析, 情報処理学会論文誌, Vol. 46, No. 3, pp. 831–844 (2005).