係り受け構造に基づく自然論理タブロー法

高尾 大樹 † 加藤 芳秀 ‡ 松原 茂樹 §
† 名古屋大学工学部電気電子・情報工学科
‡ 名古屋大学情報連携統括本部
§ 名古屋大学大学院情報科学研究科
daiki.takao@db.ss.is.nagoya-u.ac.jp

1 はじめに

自然言語処理の重要なタスクの一つに含意関係認識がある．含意関係認識とは，テキスト T が正しいときに仮説 H も正しいと推論できるか否かを判定するタスクであり，質問応答や情報検索，テキスト要約など幅広い分野への応用が期待されている．

含意関係認識へのアプローチとして，文の表層的な情報に基づき機械学習する手法が多く報告されている．しかし，これらの手法は量化や様相性といった文の意味情報を考慮した推論が難しい．

これに対して，Mineshima et al.[3] は，日本語文の含意関係認識に対する論理的アプローチとして，日本語 CCG パーザの導出木を高階論理式に変換して推論を行う手法を提案している．この手法は，日本語の意味的な現象に基づく含意関係に焦点を当てた日本語意味論テストセットである JSeM[5] に対して，高い正答率を実現している．

別のアプローチとして，自然論理を用いて推論を行う MacCartney and Manning[2] の手法がある．自然論理とは，自然言語のレベルで一方のテキストからもう一方のテキストを推論するための論理体系であり，これを用いることで，テキストを述語論理などの形式論理の論理式に変換することなく，単調性などを考慮した推論を自然な形で行うことができるという利点がある．

しかし，自然論理を用いた手法の多くは英語を対象としており，日本語を扱う効果的な方法は知られていない．日本語に適用した例として，増田ら [6] があるが，文と文の間の含意関係しか推論できないため，テキストが複数の文からなる場合に適用できないという問題がある．

そこで本稿では，推論手法としてタブロー法を組み合わせることで，自然論理を日本語の含意関係認識に適用する手法を提案する．本手法では，係り受け構造に対するタブロー法の推論規則により，接続詞，否定，一般化量化子を扱うことができる．

2 自然論理に対するタブロー法

Muskens[4] は自然論理に対して，証明技法の一つであるタブロー法を適用することで，テキストを論理式に変換することなく推論する手法を提案している．具体的には，自然言語文を Lambda Logical Forms(LLFs) と呼ばれる，構文構造に類似した形式で表現し，LLFs に対する推論規則を定義し，それを用いて含意関係などを推論するというものである．自然論理を用いることにより，文の意味と表層表現を同時に扱うことが可能となる他，その表現能力の高さから，一般化量化子や単調性を考慮した推論も可能となる．

また，Abzianidze[1] は英語 CCG パーザの導出木から LLFs への具体的な変換方法と新たな推論規則を提案している．実際に推論システムを実装しており，テストセットを用いた実験においても，高い正答率を示している．

3 提案手法

2 節で述べた手法における推論規則は全て英語を前提としたものであり，そのまま日本語に適用することはできない．そこで本節では，日本語に対応した自然論理タブロー法を提案する．本手法では，日本語文を係り受け構造で表現し，そのパターンに基づき推論する．本手法の全体の流れは次の通りである．

前処理 入力されたテキスト T と仮説 H に対して形態素解析や係り受け解析をはじめとする前処理を行う．

含意関係の推論 得られた係り受け構造に対して，3.2.2 節で提案する推論規則を用いてタブロー法による推論を行う．

3.1 前処理

前処理としてまず，形態素解析と係り受け解析により得られた係り受け構造を，形態素を基本単位とした係り受け構造に変換する．この際，元々一つの文節に含まれていた形態素は直後の形態素に係るものとして扱う．以下では，この基本単位をチャンクと呼ぶ．

次に，それぞれのチャンクに対して推論を行うために必要となる情報を付与する．具体的には，主属性として体言や用言といった形態素の種類を，副属性として活用形や格情報などを割り当てる．つまり，格助詞は一つのチャンクとして係り受け構造に存在するのではなく，あるチャンクの副情報として扱うこととなる．また，助詞の「は」は「ハ格」の格助詞のように扱う．

最後に，量化子が省略されているチャンクに対して，省略された量化子を表す ϕ を係り元のチャンクとして加える．本手法では，何らかの格情報が割り当てられているが係り元に量化子が存在しない体言のチャンクを，量化子が省略されているとみなす．

例として,「ほとんどの大きな犬は黒い」という文の係り受け構造と，それに対して前処理を行った結果を以下に示す．

$$((\text{ほとんどの}\ (\text{大きな}\ \ \text{犬は}))\ \text{黒い})$$
$$((\text{ほとんど}_{\text{の}}\ (\text{大きな}\ \ \text{犬}_{\text{は}}))\ \text{黒い}_{\text{終}})$$

ここで，$(D\ H)$ は係り受け構造 D の主辞が係り受け構造 H の主辞へ係るような係り受け構造を表す．つまり,「ほとんど$_{\text{の}}$」や「大きな」は「犬$_{\text{は}}$」に係っている．また，下付き文字の「は」や「終」は副属性の「ハ格」や「終止形」をそれぞれ表している．

3.2 自然論理に対するタブロー法

本節では，係り受け構造に基づくタブロー法を提案する．まず，タブロー法の基本的な考え方について述べ，次に係り受け構造に対するタブロー法の推論規則を提案する．

3.2.1 タブロー法

タブロー法とは，証明したい文や論理式の反例が充足不能であることを証明することにより，間接的に目的の文や論理式を証明する反駁法の一種である．含意関係を証明する場合は，テキスト T を真，仮説 H を偽として充足不能であることが証明できるかを見ればよい．

タブロー法による証明の基本的な操作は，与えられた文や論理式の集合 S に対して定められた推論規則を適用し，新たな文や論理式を生成して下に書き加えることである．このとき生成される文や論理式は元の文や論理式に含意されるものに限られる．また，複数の文や論理式のうち少なくとも一つは含意されるような場合には，枝分かれをした先にそれぞれの文や論理式を生成する．この処理を繰り返すことで得られる木構造は，初めの集合 S からの帰結の系列を表したものであり，これをタブローと呼ぶ．

タブローにおいて，S から葉までの系列を枝といい，ある枝に矛盾が生じた場合は，その枝の先に $\times$ をつけて閉じる．そして，全ての枝が閉じているとき，S は充足不能であることが証明される．

$$\mathrm{T}\colon ((\text{ほとんど}_{*}\ N_x) V_{\text{終}})$$
$$X : ((\text{ほとんど}_{*}\ N_x) V'_{\text{終}})$$
$$|$$
$$\mathrm{T}\colon (c_{\text{は}}\ V_{\text{終}})$$
$$X : (c_{\text{は}}\ V'_{\text{終}})$$
$$\mathrm{T}\colon ((c_{\text{は}}\ N_{-})\ \text{だ}_{\text{終}})$$
$$\text{ただし}, c\ \text{は未割り当て}$$

図 1: 「ほとんど」に関する規則

3.2.2 推論規則

以下では提案する推論規則の一つとして，文献 [1] を参考にして設計した量化子「ほとんど」に関する規則について説明する．まず具体例を用いて，その推論規則がどのような考え方に基づいているかを説明し，続いて，それを一般化したタブロー法の推論規則について述べる．

「ほとんどの大きな犬は黒い」が真であるとき,「大きな犬」である実体の集合を S，その要素数を $|S|$ とすると,「ほとんどの」の意味を考えれば，S において黒いという性質を持つ実体の数は $|S|/2$ 以上である．このとき,「ほとんどの大きな犬は賢い」も真である状況について考えると，賢いという性質を持つ S 中の実体の数も $|S|/2$ 以上となる．したがって，黒いという

T: A F: B × ただし, $A = B$[1]	T: ((全て$_*$ N_x)$V_{\text{終}}$) F: (($c_{\text{は}}$ N_-) だ$_{\text{終}}$) T: (c_x $V_{\text{終}}$) c は任意	T: (($c_{\text{は}}$(M_* N_-)) だ$_{\text{終}}$) T: (($c_{\text{は}}$ N_-) だ$_{\text{終}}$) ただし, M は *subsective*[2]
X : ((その$_*$ N_x)$V_{\text{終}}$) T: (($c_{\text{は}}$ N_-) だ$_{\text{終}}$) X : (c_x $V_{\text{終}}$) ただし, c は未割り当て	F: (($c_{\text{は}}$(M_* N_-)) だ$_{\text{終}}$) F: (($c_{\text{は}}$ N_-) だ$_{\text{終}}$) F: ($c_{\text{は}}$ $M_{\text{終}}$) T: (($c_{\text{は}}$ N_-) だ$_{\text{終}}$) ただし, M は *intersective*[3]	T: ((ある$_*$ N_x)$V_{\text{終}}$) T: (($c_{\text{は}}$ N_-) だ$_{\text{終}}$) T: (c_x $V_{\text{終}}$) ただし, c は未割り当て
X : ((ϕ_- N_x)V_*) X : ((ある$_-$ N_x)V_*) ただし, $x \neq$「は」	X : ((ϕ_- $N_{\text{は}}$)V_*) X : ((全て$_{\text{の}}$ $N_{\text{は}}$)V_*)	T: (($V_{\text{仮}}$ ならば)V'_*) F: $V_{\text{終}}$ T: V'_*

図 2: 推論規則の例

性質と賢いという性質をともに持つ実体が S 中に少なくとも一つ存在すると推論できる．

一方,「ほとんどの大きな犬は賢い」が偽であれば，S において黒いという性質を持つ実体の数は $|S|/2$ 以上であるものの，賢いという性質を持つ実体の数は $|S|/2$ に満たない．したがって，黒いという性質を持つが賢いという性質を持たない実体が S 中に少なくとも一つ存在すると推論できる．

以上のような推論をタブロー法の推論規則として一般化したものが図 1 である．今回提案するタブロー法におけるタブローの各ノードは真偽値 (T は真，F は偽を意味する) と係り受け構造の対からなり，タブローの枝上に規則の上部にマッチするノードが存在するとき，規則の下部を生成する．この規則において X は任意の真偽値であり，c は実体を表す定項，ほとんど$_*$ は量化子「ほとんど」と同じ意味を持つチャンク，N 及び V は主属性が体言や用言であるチャンクを主辞とする係り受け構造，x は任意の副属性をそれぞれ指している．上述の例では，N に「(大きな 犬)」，x に「は」，V に「黒い」，V' に「賢い」がそれぞれマッチングするので,「T: ($c_{\text{は}}$ 黒い$_{\text{終}}$)」,「F: ($c_{\text{は}}$ 賢い$_{\text{終}}$)」，「T : (($c_{\text{は}}$ (大きな 犬)) だ)」の 3 つのノードが生成される．

このように各形態素の表す意味を考慮し，ある係り受け構造から生成できる係り受け構造の組み合わせを，文献 [1][4] 等を参考にしながら，タブロー法の推論規則として設計した．図 2 に一例を示す．一般化量化子や接続詞，否定などに関する規則を合計 29 個設計した．

4 適用例

例として，T が「ほとんどの大きな犬は黒い」と「全ての黒い犬はとても賢い」，H が「ほとんどの大きな犬はとても賢い」である場合に，本手法により含意関係を推論する流れを説明する．

まず初めに，形態素・係り受け解析を行って得られた係り受け構造に対して前処理を行うと以下のような係り受け構造が得られる．なお，今後の処理に関係の無い情報は省略している．

((ほとんど$_{\text{の}}$ (大きな 犬$_{\text{は}}$)) 黒い$_{\text{終}}$)

((全て$_{\text{の}}$ (黒い 犬$_{\text{は}}$)) (とても 賢い$_{\text{終}}$))

((ほとんど$_{\text{の}}$ (大きな 犬$_{\text{は}}$)) (とても 賢い$_{\text{終}}$))

次に，T が H を含意するかを推論するために，T の文の係り受け構造に真を，H の文の係り受け構造に偽を

[1] 文献 [1] における論理的な順序関係 $\leq$ を日本語において定義できれば，$A \leq B$ とすることができる．

[2] 性質 M を持つ実体の集合を $[\![M]\!]$ とすると，$[\![(M\ N)]\!] \subseteq [\![N]\!]$ が成立するとき，M は *subsective* という．

[3] 性質 M を持つ実体の集合を $[\![M]\!]$ とすると，$[\![(M\ N)]\!] = [\![M]\!] \cap [\![N]\!]$ が成立するとき，M は *intersective* という．

$\boxed{1}$ T: ((ほとんど$_{の}$ (大きな 犬$_{は}$)) 黒い$_{終}$)

$\boxed{2}$ T: ((全て$_{の}$ (黒い 犬$_{は}$)) (とても 賢い$_{終}$))

$\boxed{3}$ F: ((ほとんど$_{の}$ (大きな 犬$_{は}$)) (とても 賢い$_{終}$))

$\boxed{4}$ T: ($c_{は}$ 黒い$_{終}$)

$\boxed{5}$ F: ($c_{は}$ (とても 賢い$_{終}$))

$\boxed{6}$ T: (($c_{は}$ (大きな 犬$_{-}$)) だ$_{終}$)

$\boxed{7}$ T: (($c_{は}$ 犬$_{-}$) だ$_{終}$)

$\boxed{8}$ F: (($c_{は}$ (黒い 犬$_{-}$)) だ$_{終}$) $\boxed{13}$ T: ($c_{は}$ (とても 賢い$_{終}$))

$\boxed{14}$ ×

$\boxed{9}$ F: (($c_{は}$ 犬$_{-}$) だ$_{終}$) $\boxed{11}$ F: ($c_{は}$ 黒い$_{終}$)

$\boxed{10}$ × $\boxed{12}$ ×

図 3: 入力例に対するタブロー

割り当てて推論を行う．推論の結果得られるタブローを図 3 に示す．

$\boxed{1}$と$\boxed{3}$のノードに対して図 1 の規則を適用することで，$\boxed{4}$，$\boxed{5}$，$\boxed{6}$のノードが生成される．次に，$\boxed{6}$に対して図 2 の subsective である修飾語に関する規則を適用することで，$\boxed{7}$が生成される．さらに図 2 中の規則を適用していくと，最終的に$\boxed{7}$と$\boxed{9}$，$\boxed{4}$と$\boxed{11}$，$\boxed{5}$と$\boxed{13}$それぞれに対して図 2 の枝を閉じる規則を適用することで，タブローの全ての枝が閉じられる．よって，T が真，H が偽であるときに矛盾が生じることから，T と H に含意関係が成立するといえる．

5 まとめ

本稿では，日本語に対応した自然論理タブロー法による推論手法とその適用例を示した．現在，本稿で提案したタブロー法を実装中である．今後，テストセットを用いた実験により本手法の有用性を確認する予定である．

参考文献

[1] Lasha Abzianidze. Natural Solution to FraCaS Entailment Problems. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics (*SEM 2016)*, pp. 64–74, 2016.

[2] Bill MacCartney and Christopher D. Manning. Modeling Semantic Containment and Exclusion in Natural Language Inference. In *Proceedings of the 22nd International Conference on Computational Linguistics, Volume 1*, pp. 521–528, 2008.

[3] Koji Mineshima, Ribeka Tanaka, Pascual Martınez-Gómez, Yusuke Miyao, and Daisuke Bekki. Building Compositional Semantics and Higher-Order Inference System for a Wide-Coverage Japanese CCG Parser. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2236–2242, 2016.

[4] Reinhard Muskens. An Analytic Tableau System for Natural Logic. In *Logic, Language and Meaning*, pp. 104–113. Springer, 2010.

[5] 川添愛, 田中リベカ, 峯島宏次, 戸次大介. 形式意味論に基づく含意関係テストセット構築の方法論. 第 29 回 人工知能学会全国大会論文集, 2015.

[6] 増田涼良, 杉本徹. Natural Logic を用いた日本語テキストの含意関係認識. 第 26 回 人工知能学会全国大会論文集, 2012.