ニューラルモデルとタブロー法を用いた自然言語推論のモデル理論的な定式化

佐治 礼仁 [1]　加藤 芳秀 [2]　松原 茂樹 [1,2]
[1] 名古屋大学大学院情報学研究科　[2] 名古屋大学情報連携推進本部
saji.ayahito.y7@s.mail.nagoya-u.ac.jp

概要

自然言語推論とは，2 つのテキストの間に成立する推論的関係を同定するタスクである．Saji ら [1] は，ニューラルモデルに基づく自然言語推論システムと記号操作的な証明手法であるタブロー法を組み合わせた手法を提案したが，その理論的限界は明らかではない．本論文では，Saji らの手法をモデル理論的に定式化し，その定式化に基づいて，手法の性質とその限界を明らかにする．

1 はじめに

自然言語推論とは，2 つのテキスト（一方を**前提**，他方を**仮説**と呼ぶ）の間に成り立つ推論的関係を同定するタスクである．論理的知識や常識的知識を用いて前提から仮説が導出可能である場合は**含意**，前提と仮説が両立しえない場合は**矛盾**，そのいずれでもない場合は**中立**と判定する．質問応答や情報検索，テキスト要約などの幅広い分野での利用が期待されている．

近年，ニューラルモデルに基づく手法が自然言語推論において用いられている（例えば，[2]）．ニューラルモデルは単純でありながら，SNLI コーパス [3] や MNLI コーパス [4] を用いた実験において高い正解率を達成している．しかし一方で，この手法には，判定結果に至る過程や理由を説明する能力を持たないという問題がある．ニューラルモデルの内部はブラックボックス化しており，どのような推論が行われたのかを推察することも困難である．一方，自然言語推論への記号操作的な手法（例えば，[5, 6, 7]）が提案されている．この手法は，ニューラルモデルによる手法と異なり，結果に至る推論過程が人間にとって理解可能であるという利点が挙げられる．また，この手法における記号操作は一般に，形式論理学や言語学による裏付けが与えられており，そのような推論を行った根拠を与えることができる．

Saji ら [1] はこの二つの手法を組み合わせ，ニューラル自然言語推論モデルに，記号操作的手法のような推論過程を明示化する性質を付加する手法を提案している．この手法は，形式論理の証明法の一つであるタブロー法をその背景としているが，その理論的限界は明らかにされていなかった．

そこで本論文では，モデル理論的に Saji らの手法を定式化し，その定式化に基づいて，この手法の理論的限界を明らかにする．具体的には Saji らの手法について，ある種の健全性が成立する一方で，完全性が成立しないことを示す．

2 ニューラルモデルとタブロー法を用いた自然言語推論

2.1 タブロー法の概略

本節では，Saji らのタブロー法について説明する．このタブロー法は，自然言語文と真偽値の対の集合が与えられたとき，その集合中のすべての文にその対となる真偽値を割り当てることが可能か否かを証明する．自然言語推論とタブロー法の関係は次のように整理されている．[1)]

- 前提，仮説が共に真となるような真偽値割り当てが存在しないことが証明されるとき，それは前提と仮説が矛盾の関係にあることを意味する．
- 前提が真，仮説が偽となるような真偽値割り当てが存在しないことが証明されるとき，前提が真であるならば仮説も真となるのは必然である，すなわち，前提が仮説を含意することを意味する．
- どちらの証明もできない場合は，前提と仮説は中立の関係にあることを意味する．

1）この整理は不十分であることが 3 節において示される．

タブロー法では，与えられた文，真偽値の対の集合をもとに，タブローと呼ばれる木構造を構成する．タブローのノードのラベルは，3 つ組 $(s, X) : Y$ であり，**エントリ**と呼ばれる．これは，文 s が真偽値として X を取らなければならないという制約を示しており，Y は，このエントリに対して後述するタブロー規則が適用されたか否かを表すフラグである．真偽値は，T か F のいずれかであり，それぞれ真と偽を表す．フラグは，0 か 1 のいずれかであり，0 はタブロー規則が未適用，1 は適用済みであることを表す．初期タブローは，与えられた自然言語文と真偽値の対に対して，規則が適用されていないことを示す値 0 をフラグとして付加したエントリから構成される．これにタブロー規則をエントリに対して繰り返し適用することによりタブローは導出される．タブロー規則の適用により，エントリが表現する制約は，より単純な文に関する制約へと分解され，新たなエントリとしてタブローに付け加えられる．タブローにおける枝分かれは場合分けに相当する．この分解により，推論の過程が明示化される．

タブロー規則は，以下の形式で表現される．[2)]

$$(C_{1,1} \wedge \cdots \wedge C_{1,n_1}) \vee \cdots \vee (C_{m,1} \wedge \cdots \wedge C_{m,n_m})$$

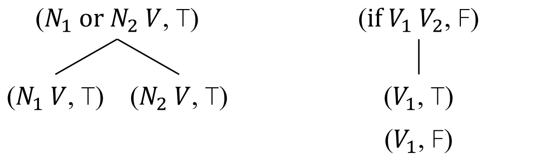
ただし，$C_{1,1}, \cdots, C_{1,n_1}, \cdots, C_{m,1}, \cdots, C_{m,n_m}$ は自然言語文と真偽値の対を入力とし，自然言語文と真偽値を出力する関数である．タブロー中に，次のようなエントリ $(s, X) : Y$ が存在するとき，それが支配するすべての葉に新たな m 個の枝 $\langle C_{1,1}(s, X) : 0, \cdots, C_{1,n_1}(s, X) : 0\rangle, \cdots, \langle C_{m,1}(s, X) : 0, \cdots, C_{m,n_m}(s, X) : 0\rangle$ を追加する．

- すべての C_{i,j_i} について，$C_{i,j_i}(s, X)$ は未定義でない．[3)]
- $Y = 0$.

タブロー規則の適用は，適用元のエントリが表現する制約を，それと等価な制約へと変換する操作と位置づけられる．例えば，(Either Smith or Anderson signed the contract, T): 0 に図 1 左の規則を適用すると，その経路の終点（タブローの葉）に 2 つのエントリ (Smith signed the contract, T): 0 と (Anderson signed the contract, T): 0 が追加される．また，導出元のエントリである (Either Smith or Anderson signed the contract, T): 0 のフラグを 1 に変更する．新たに追加

2) 具体的なタブロー規則の実現方法については，文献 [1] を参照のこと．

3) 未定義である場合は，タブロー規則が適用できないことを意味する．



選言に関する規則　　条件法に関する規則

図 1　タブロー規則の例

された 2 つのエントリにより表現される制約は，元のエントリの制約と等価である．

2.1.1　タブローの閉鎖

真偽値を割り当てることができないという事実は，経路，及びタブローの閉鎖という概念により表現される．一般のタブロー法においては，経路上に真偽値のみが異なるエントリが存在することをもって閉鎖経路を定義しているが，Saji らの手法では，これに加えてニューラル自然言語推論モデルに基づく閉鎖経路を定義している．自然言語推論モデルは，前提 p と仮説 h を入力として受け取ると，それを含意 (E)，中立 (N)，矛盾 (C) のいずれかのクラスに分類すると想定する．以下では，$L \times L$（L は文の集合）から {E, N, C} への関数を **NLI システム**と呼び，$NLI(p, h) = \mathrm{C}$ といった表記を用いる．経路上に 2 つのエントリ $(s_1, X_1) : 0, (s_2, X_2) : 0$ が存在し，次のいずれかが成り立つとき，その経路は閉じているという．以下では，これを経路の**閉鎖条件**と呼ぶ．

- $X_1 = \mathrm{T} \wedge X_2 = \mathrm{F} \wedge s_1 = s_2$
- $X_1 = \mathrm{T} \wedge X_2 = \mathrm{T} \wedge NLI(s_1, s_2) = \mathrm{C}$
- $X_1 = \mathrm{T} \wedge X_2 = \mathrm{F} \wedge NLI(s_1, s_2) = \mathrm{E}$

全ての経路が閉じているとき，タブローは閉じるという．タブローが閉じていることは，与えられたエントリの集合が表す制約を満たす真偽値割り当てが存在しないことに相当する．

推論過程の明示化という観点からは，タブロー規則をできる限り適用したタブローに対して閉鎖性を判定するのが望ましい．以下では，そのようなタブローを区別するために，初期タブローにタブロー規則を適用できなくなるまで適用を繰り返したタブローを**終端タブロー**と呼ぶことにする．

前提文を T，仮説文を F としたときの終端タブローを**含意タブロー**と呼び，上で述べた 1 番目のケースの証明に用いる．前提文と仮説文の両方を T としたときの終端タブローを**矛盾タブロー**と呼び，これは 2 番目のケースの証明に用いる．

3　モデル理論的な定式化

3.1　モデル

本節では，Saji らの手法を形式的に議論するために，まず，その基盤となるモデル理論的な文の解釈について定義する．次に，モデルに基づいて，タブロー規則，及び NLI システムを特徴付け，タブロー規則により導出された終端タブローにおける NLI システムを用いた閉鎖性の判定のもつ性質を明らかにする．

まず，モデルを定義する．

定義 1 (モデル)**.** L から $\{\mathsf{T},\mathsf{F}\}$ への関数 m をモデルと呼ぶ．$\mathcal{M}$ をすべてのモデルの集合とする．モデルの集合 $M \subseteq \mathcal{M}$ に対して，$M(s) = \{m \in M \mid m(s) = \mathsf{T}\}$ と定義する．

直観的には，モデルの集合 $M(s)$ は，文 s が成り立つ状況の集合とみなすことができる．

以下では，モデル集合が与えられたとき，それと整合性のとれた NLI システム及びタブロー規則がどのようなものであるかを定義し，整合性が取れているという条件の下で Saji らのタブロー法による証明の理論的限界を明らかにする．

定義 2. 次の 3 つの条件が全て成り立つとき，NLI システムはモデル集合 M と**整合的**であるという．

- $M(s_1) \subseteq M(s_2) \Leftrightarrow NLI(s_1, s_2) = \mathsf{E}$
- $\neg(M(s_1) \subseteq M(s_2)) \land \neg(M(s_1) \cap M(s_2) = \emptyset) \Leftrightarrow NLI(s_1, s_2) = \mathsf{N}$
- $M(s_1) \cap M(s_2) = \emptyset \Leftrightarrow NLI(s_1, s_2) = \mathsf{C}$

この定義では，含意関係をモデル集合の包含関係として，矛盾関係を互いに素な関係として捉えている．

以下では，モデルに基づいてタブローを特徴付けるが，その準備として，エントリ，タブローの経路，タブローに対応するモデルを，$M(s)$ を拡張することにより定義する．

定義 3 (エントリに対するモデル集合)**.** タブローのエントリ $e = (s, X) : Y$ に対して，$M(e)$ を次のように定義する．

$$M(e) = \begin{cases} M(s) & (X = \mathsf{T}) \\ M - M(s) & (X = \mathsf{F}) \end{cases}$$

定義 4 (タブローの経路に対するモデル集合)**.** タブローの経路 b に対して，$M(b)$ を次のように定義する．

$$M(b) = \cap_{e \in b} M(e)$$

定義 5 (タブローに対するモデル集合)**.** タブロー t に含まれる全ての経路の集合を B とする．$M(t)$ を次のように定義する．

$$M(t) = \cup_{b \in B} M(b)$$

2 節で述べたように，Saji らの手法ではタブロー規則の適用前後で，エントリが表現する制約が等価であることが想定されているが，これを以下のように定義する．

定義 6. $r = (C_{1,1} \land \cdots \land C_{1,n_1}) \lor \cdots \lor (C_{m,1} \land \cdots \land C_{1,n_m})$，$E = L \times \{\mathsf{T},\mathsf{F}\} \times \{0,1\}$，$E_r = \{e \in E \mid r$ は e に適用可能 $\}$ とする．任意の $e \in E_r$ に対して次の条件が満たされるとき，タブロー規則 r はモデル集合 M と**整合的**であるという．

$$M(e) = \Big(M\big(C_{1,1}(e)\big) \cap \cdots \cap M\big(C_{1,n_1}(e)\big)\Big) \cup \cdots \cup \Big(M\big(C_{m,1}(e)\big) \cap \cdots \cap M\big(C_{1,n_m}(e)\big)\Big)$$

また，任意の $r \in R$ が M と整合的であるとき，R は M と整合的であるという．

3.2　Saji らのタブロー法の健全性と完全性

本節では，前節で定義したモデルに基づき Saji らのタブロー法における健全性と完全性を定義する．また，健全性について証明し，完全性について反例を示す．

ここで定義する健全性は，含意（矛盾）タブローが閉じるならば，前提と仮説に対するモデル集合は，包含（互いに素な）関係にあるという性質である．これを前節で導入したモデルに基づき定式化すると，次のように定義できる．

定義 7 (含意タブロー・矛盾タブローの健全性)**.** M をモデル集合，R を M と整合的なタブロー規則の集合，NLI を M と整合的である NLI システムであるとする．任意の前提文 p と仮説文 h について，次の条件が成り立つならば，R と NLI による閉鎖性判定は，モデル集合 M において**健全**であるという．

- R によって構成された含意タブローが NLI によって閉じると判定されるならば，$M(p) \subseteq M(h)$ が成り立つ．
- R によって構成された矛盾タブローが NLI によって閉じると判定されるならば，$M(p) \cap M(h) = \emptyset$ が成り立つ．

一方で，完全性は次のように定義する．

定義 8 (含意タブロー・矛盾タブローの完全性)**.** M をモデル集合，R を M と整合的なタブロー規則の集合，NLI を M と整合的である NLI システムであるとする．任意の前提文 p と仮説文 h について，次の条件が成り立つならば，R と NLI による閉鎖性判定は，モデル集合 M において**完全**であるという．

- $M(p) \subseteq M(h)$ が成り立つなら，R によって構成された含意タブローが NLI によって閉じると判定される．
- $M(p) \cap M(h) = \emptyset$ が成り立つならば，R によって構成された矛盾タブローが NLI によって閉じると判定される．

3.2.1 健全性の証明

本節では，まず健全性を証明するための補助定理を導入し，それにも基づき健全性を証明する．

補助定理 1. R が M と整合的であるならば，任意の $r \in R$ の適用前のエントリの集合に対応するモデル集合と，その r の適用後のエントリの集合に対応するモデル集合が等しい．

補助定理 1 は定義 6 から明らかである．

補助定理 2. t_{start} を初期タブローとし，t_{trm} を t_{start} から導かれる終端タブローとする．R が M と整合的であるならば，$M(t_{\text{start}}) = M(t_{\text{trm}})$ である．

補助定理 2 は補助定理 1 から明らかである．含意タブローの健全性の証明は次の通りである．R によって構成された含意タブロー t_{trm} が NLI によって閉じていると判定されるとする．このとき，t_{trm} のすべての経路 b に閉鎖条件のいずれかを満たす 2 つのエントリ $(s_1, X_1) : 0, (s_2, X_2) : 0$ が存在する．

- 1 番目の閉鎖条件を満たすならば，定義 1 より $s_1 = s_2$．したがって，$M\big((s_1, \mathsf{T})\big) \cap M\big((s_2, \mathsf{F})\big) = M\big((s_1, \mathsf{T})\big) \cap M\big((s_1, \mathsf{F})\big) = \emptyset$.
- 2 番目の閉鎖条件を満たすならば，定義 2 より $NLI(s_1, s_2) = \mathsf{C}$ であるが，NLI は M と整合的であるので $M(s_1) \cap M(s_2) = \emptyset$．よって，定義 3 より $M\big((s_1, \mathsf{T})\big) \cap M\big((s_2, \mathsf{T})\big) = \emptyset$.
- 3 番目の閉鎖条件を満たすならば，定義 2 より $NLI(s_1, s_2) = \mathsf{E}$ であるが，NLI は M と整合的であるので $M(s_1) \subseteq M(s_2)$．よって，定義 3 より $M\big((s_1, \mathsf{T})\big) \subseteq M\big((s_2, \mathsf{T})\big)$ すなわち，$M\big((s_1, \mathsf{T})\big) \cap M\big((s_2, \mathsf{F})\big) = \emptyset$.

以上より，$M\big((s_1, X_1)\big) \cap M\big((s_2, X_2)\big) = \emptyset$．ゆえに，定義 4 より $M(b) = \emptyset$ となる．したがって，定義 5 から $M(t_{\text{trm}}) = \emptyset$ である．t_{start} を t_{trm} の初期タブローとすると，補助定理 2 から $M(t_{\text{start}}) = \emptyset$ である．初期タブローは $(p, \mathsf{T}) : 0, (h, \mathsf{T}) : 0$ から構成されるため，定義 4, 5 から $M\big((p, \mathsf{T})\big) \cap M\big((h, \mathsf{F})\big) = \emptyset$．したがって，$M\big((p, \mathsf{T})\big) \subseteq M\big((h, \mathsf{T})\big)$．定義 3 から $M(p) \subseteq M(h)$ である．以上より題意は満たされる．矛盾タブローの健全性も同様に証明できる．

$\boxed{1}$ (Smith and Anderson didn't go out, T): 1

$\boxed{2}$ (Smith and Anderson were home, F): 1

$\boxed{3}$ (Smith or Anderson went out, F): 1

$\boxed{4}$ (Smith went out, F): 0

$\boxed{5}$ (Anderson went out, F): 0

$\boxed{6}$ (Smith was home, F): 0 $\boxed{7}$ (Anderson was home, F): 0

図 2 タブローの完全性の反例

3.2.2 完全性の反例

本節では，含意タブローの完全性の反例を示す．

図 2 に示す含意タブローにおいて，タブロー規則が未適用であるエントリ $\boxed{4}\boxed{5}\boxed{6}\boxed{7}$ の真偽値はいずれもが F である．閉鎖条件においては，いずれかのエントリの真偽値が T でなければならない．このため，この含意タブローは，どのような NLI を用いても閉じていると判定されることはない．この例に対して M を用意し，それと整合的で $NLI(\boxed{1}, \boxed{2}) = \mathsf{E}$（$\Leftrightarrow M(\boxed{1}) \subseteq M(\boxed{2})$）であるような NLI は容易に構成できるが，この M 及び NLI において含意タブローは閉じると判定されない．すなわち，完全性が成立しない．矛盾タブローについても同様の反例が存在する．つまり，完全性は一般に成立しない．

4 おわりに

本論文は，Saji らの手法をモデル理論的に定式化し，その定式化に基づいて，その性質を明らかにした．一般に，健全性は成り立つが完全性は成り立たないことを示した．現在，タブロー規則集合が特定の条件を満たすとき，完全性が成り立つようなケースを確認しているが，今後の課題としては，そのような条件がいかなるものであるかの分析が挙げられる．

参考文献

[1] Ayahito Saji, Daiki Takao, Yoshihide Kato, and Shigeki Matsubara. Natural language inference using neural network and tableau method. In **Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation**, pp. 406–414, Shanghai, China, November 2021. Association for Computational Linguistics.

[2] Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, Hui Jiang, and Diana Inkpen. Enhanced LSTM for natural language inference. In **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1657–1668, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[3] Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. A large annotated corpus for learning natural language inference. In **Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing**, pp. 632–642, Lisbon, Portugal, September 2015. Association for Computational Linguistics.

[4] Adina Williams, Nikita Nangia, and Samuel Bowman. A broad-coverage challenge corpus for sentence understanding through inference. 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. 1112–1122, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[5] Bill MacCartney and Christopher D. Manning. Natural logic for textual inference. In **Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing**, pp. 193–200, Prague, June 2007. Association for Computational Linguistics.

[6] Lasha Abzianidze. A tableau prover for natural logic and language. In **Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing**, pp. 2492–2502, Lisbon, Portugal, September 2015. Association for Computational Linguistics.

[7] Lasha Abzianidze. LangPro: Natural language theorem prover. In **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 115–120, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.