

テキスト推論でセンター試験の歴史問題を解く

Answering Center-exam Questions on History by Textual Inference

田 然*¹ 宮尾 祐介*²
Ran Tian Yusuke Miyao

*¹国立情報学研究所 *²国立情報学研究所
National Institute of Informatics National Institute of Informatics

A great portion of Center-exam questions on history consists of fact validation, the task to decide whether a natural language sentence states a historical fact. The distinction between historical facts and non-facts, especially in exam questions, is thought to be clear. However, to clarify the criterion and to realize it in a computer system, is a challenging topic in artificial intelligence and natural language processing. In this paper, we propose a semantic representation suited for natural language inference, which can be used to model such criterions; and we discuss some preliminary experiments on answering fact validation questions by clear criteria.

1. はじめに

国立情報学研究所が推進している「人工頭脳プロジェクト—ロボットは東大に入れるか」は、人工知能による大学入試突破を目標としており、その中で世界史・日本史といった歴史科目のセンター試験では、自然言語で書かれている文が史実であるかどうかを判断する問題が大部分を占める。入試問題の性質上、「史実である文」と「史実ではない文」との間に明確な線引きがあると考えられている。この判定基準を明確にし、計算機によってそれをどれほど実現できるかは、人工知能および自然言語処理における興味深い問題である。

史実である文とそうでない文を区別する基準を与えるオントロジーが提案された [川添 12] が、自然言語で書かれている文をどのようにオントロジーと結びつけ、オントロジーが定める推論規則をどのように実装していくのかは、いまだ多くの問題が残る。本稿では、自然言語の推論に適した意味表現を提案し、これにより自然言語文の正誤を判定することができることを示す。また、それを実装した推論エンジンの性能と、明確な推論規則を用いてセンター試験歴史科目に解答する予備実験について考察する。

2. 文の正誤を判定する推論システム

明確な推論規則を適用できるように、自然言語文を論理形式に変換し自動証明を行う研究は以前から行われてきた [Moldovan 03, Bos 06, Raina 05, Beltagy 13]。ほとんどのシステムは図 1 に示す構成をなすが、使用する文法や形式意味論、知識の生成手段などは様々である。論理表現として、主に一階述語論理 (FOL) が使われてきたが、FOL をフルに扱う論理推論は重いため、それが障害になる場合もある。例えば [Beltagy 13] は、論理式の中の全称量子子を全て存在量子子に変換するなどの妥協手段を使わざるを得なかった。

一階述語論理よりも限定された論理体系で自然言語の推論を実現する試みも提案された。例えば [MacCartney 08] は Natural Logic を用いた手法で効率的推論を実現し、[Schoenmackers 08] では単純な意味表現を使った論理推論を大規模なウェブデータに適用する場合を考察した。しかし、これらの論理体系で表現できる意味は非常に限定的で、自然言語

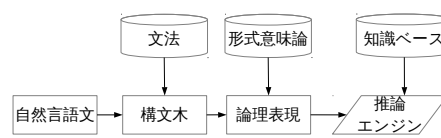


図 1: テキスト推論の模式図

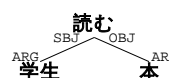


図 2: 「学生が本を読む」の DCS 木

に見られる推論現象 (共参照や時間関係など) の多くは対象外である。

我々のシステムは、構文木として係り受け木を用いる。次に、DCS (Dependency-based Compositional Semantics) [Liang 11] の形式意味論に基づいて文の論理表現を得る。DCS は、係り受け木に近い構造を持つ DCS 木 (図 2) で文の意味を厳密に表す枠組みである。DCS 木の定める意味に対して効率的な推論を行うために、我々は新しい論理表現、外延の抽象表現を提案する (§3.)。そして、外延の抽象表現に基づく推論エンジンを実装した。推論エンジンを使って、辞書から得られる類義語・反義語知識、教科書に書かれた歴史知識 (図 1 のように外延の抽象表現に変換する)、そして動的に生成される知識 (§4.1) から、証明できる文を「史実である正しい文」として認識する。

3. 外延の抽象表現を用いた論理推論

DCS は、自然言語文を関係データベースのクエリに変換する枠組みとして提案された [Liang 11]。図 2 のように、「学生が本を読む」という文は係り受け木の形で表され、木の各辺には SBJ や OBJ のような意味役割が付与される。このような木 (DCS 木) が与えられると、データベースに対する検索処理が一つ定義される。例えば、表 1 のデータベースに対し、図 2 の DCS 木は、「読む」表の中から SBJ が「学生」、OBJ が「本」に入るレコード (「花子が平家物語を読む」など) を取り出す。その結果、「学生が本を読む」が表すレコードの集合 (外延) が

学生		本		読む	
ARG		ARG		SBJ	OBJ
太郎		平家物語		社員 A	日本経済新聞
花子		古事記		花子	平家物語
のび太		...		ドラえもん	小学五年生
...			

表 1: 学生、本、読む のデータベース

得られる。全称量子化などより複雑な意味も、DCS 木に記号を追加し、データベースに対する適切な操作を対応させることで表現できる。例えば図 3 では、**辺教科書 ARG-OBJ 読む**にある「C」記号が全称量化を表し、これは除算 q_C^{OBJ} (§3.1 参照) に対応する。

[Liang 11] ではデータベースに対する検索命令として DCS の意味を定義しているが、関係代数の演算子を用いてデータベースの検索命令を抽象化すると、DCS 木の意味を表す**外延の抽象表現**を定義することができる [田 14]。例えば「学生が本を読む」は、以下のように表す。

$$\text{読む} \cap (\text{学生}_{SBJ} \times \text{本}_{OBJ}) \quad (1)$$

ここで \cap と \times は関係代数の演算子で、それぞれ交わりと直積を表す。(1) の式は、学生と本の集合 (表 1 における「学生」と「本」の表にそれぞれ対応) の直積を作り、それと**読む**の集合の共通部分をとることを表す。これを表 1 のデータベースに当てはめれば、DCS 木で計算したものと同一集合を取り出すことができる。ただし、(1) は具体的なデータベースに依存しない数式であり、直接推論の対象とすることができる論理表現であることが重要である。この表現を使った推論によって、例えば「学生に読まれる本」の意味を表すレコードの集合は、(どんなデータベースにおいても)「誰かに読まれる出版物」を表すレコード集合の部分集合であることを示すことができる。

3.1 外延の抽象表現

形式的に、外延の抽象表現とは、下記の定数

- W : 全てのエンティティを含む普遍集合。
- 内容語が定める集合 ($\text{読む} = \{(x, y) \mid \text{読む}(x, y)\}$)。

に次の関数

- \times : 二つの集合の直積。
- \cap : 二つの集合の交わり。
- π_r : 射影 ($\pi_{OBJ}(\text{読む}) = \{y \mid \exists x; \text{読む}(x, y)\}$)。
- ι_r : ラベル付け ($\iota_{OBJ}(\text{本}) = \text{本}_{OBJ}$)。
- q_C^r : 除算。 $q_C^r(A, B)$ は $B_r \times q_C^r(A, B) \subset A$ を満たす最大の集合。
- $q_{||}^r$: 同じように $q_{||}^r(A, B)$ は $B_r \times q_{||}^r(A, B) \parallel A$ を満たす最大の集合。

を適用して得られる全ての項を指す。

3.2 言明

DCS 木の意味は外延の抽象表現で表されるが、平叙文の意味は外延の抽象表現に対する言明で表される。言明は外延の抽象表現の満たす関係であり、我々は以下三つの集合間関係を考える。

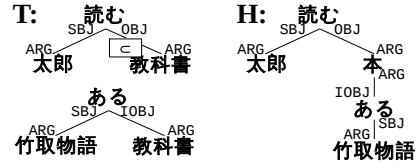


図 3: 「太郎は全ての教科書を読む」(左上)、「竹取物語は教科書にある」(左下)、「太郎は竹取物語のある本を読む」(右)の DCS 木

1. $W \neq \emptyset$
2. $\text{type}(A) = (r_1, r_2) \Rightarrow A \subset W_{r_1} \times W_{r_2}$
3. $A \subset A$
4. $A \subset B \ \& \ B \subset A \Rightarrow A = B$
5. $A \subset B \ \& \ B \subset C \Rightarrow A \subset C$
6. $A \cap B \subset A$
7. $(C \subset A \ \& \ C \subset B) \Rightarrow C \subset A \cap B$
8. $A \subset B \ \& \ A \neq \emptyset \Rightarrow B \neq \emptyset$
9. $A \parallel B \Rightarrow B \parallel A$
10. $A \parallel B \ \& \ C \subset A \Rightarrow C \parallel B$
11. $\pi_r(A) \parallel \pi_r(B) \ \& \ \text{type}(A) = \text{type}(B) \Rightarrow A \parallel B$
12. $A \parallel A \Rightarrow \pi_r(A) \parallel \pi_r(A)$
13. $\pi_r(A) \neq \emptyset \Leftrightarrow A \neq \emptyset$
14. $A \subset B \Rightarrow \pi_r(A) \subset \pi_r(B)$
15. $\pi_{r_1}(\pi_{r_1, r_2}(A)) = \pi_{r_1}(A)$
16. $\text{type}(A) = (r_1, r_2) \ \& \ \pi_{r_1}(A) \subset B \Rightarrow A \subset B_{r_1} \times W_{r_2}$
17. $A_{r_1} \times B_{r_2} \neq \emptyset \Leftrightarrow A \neq \emptyset \ \& \ B \neq \emptyset$
18. $A \subset B \ \& \ C \subset D \Rightarrow A_{r_1} \times C_{r_2} \subset B_{r_1} \times D_{r_2}$
19. $(A_{r_1} \times B_{r_2}) \times C_{r_3} = A_{r_1} \times (B_{r_2} \times C_{r_3})$
20. $B \neq \emptyset \Rightarrow \pi_{r_1}(A_{r_1} \times B_{r_2}) = A$
21. $(A_{r_1} \times B_{r_2}) \cap (C_{r_1} \times D_{r_2}) = (A \cap C)_{r_1} \times (B \cap D)_{r_2}$
22. $\pi_{r_1}((A_{r_1} \times B_{r_2}) \cap C) = A \cap \pi_{r_1}((W_{r_1} \times B_{r_2}) \cap C)$
23. $B_r \times q_C^r(A, B) \subset A$
24. $B \subset C \ \& \ C_r \times D \subset A \Rightarrow D \subset q_C^r(A, B)$
25. $B_r \times q_{||}^r(A, B) \parallel A$
26. $B \subset C \ \& \ C_r \times D \parallel A \Rightarrow D \subset q_{||}^r(A, B)$
27. $q_{||}^r(A, B) \parallel \pi_{r_2}((B_{r_1} \times W_{r_2}) \cap A)$
28. $C \parallel \pi_{r_2}((B_{r_1} \times W_{r_2}) \cap A) \Rightarrow C \subset q_{||}^r(A, B)$
29. $A \parallel A \ \& \ A \neq \emptyset \Rightarrow \perp$

表 2: 公理のリスト

- 非空 $A \neq \emptyset$: A が空集合でないことを表す。
- 包含 $A \subset B$: A が B に包含されることを表す。
- 排他 $A \parallel B$: A と B の交わりが空であることを表す。

直感的には、これらは妥当性、含意、矛盾にそれぞれ対応する。

3.3 自然言語文から言明へ

外延の抽象表現は DCS 木の定めた検索命令の抽象化なので、DCS 木から計算することができる。計算手続きは基本的に [Liang 11] とパラレルであるが、詳細は省略する。DCS 木は、文の係り受け木からルール変換で得られる。

3.4 論理推論

自然言語文の意味を抽象表現の言明で表すと、テキスト間の推論は言明間の論理関係に帰着される。言明は、FOL 式に変換することもできる (例えば、 $\text{読む} \cap (\text{学生}_{SBJ} \times \text{本}_{OBJ}) \neq \emptyset$ が $\exists x, y; \text{student}(x) \ \& \ \text{read}(x, y) \ \& \ \text{book}(x)$ に同値である) が、我々は抽象表現を直接扱う推論エンジンを構築した。これによって、「本」などの集合は述語ではなく定数になり、関係代数の演算子 \cap 、 \times 、 π などは関数、外延の抽象表現は論理式ではなく項になる。そして $\text{読む} \cap (\text{学生}_{SBJ} \times \text{本}_{OBJ}) \neq \emptyset$ のような言明は全て原子文になるため、効率的な推論が可能となる。

抽象表現を直接扱うためには、その適切な公理系を設計する必要がある。我々が実装した公理を表 2 に示す。これらは抽象表現や言明間の論理関係を完備に記述する公理系ではなく、自然言語における大部分の推論を可能にし、かつ計算機で高速に処理できるように設計されている。

公理系の重要な特徴として、全ての公理がホーン節であることが挙げられる。ホーン節については、高速に自動推論を

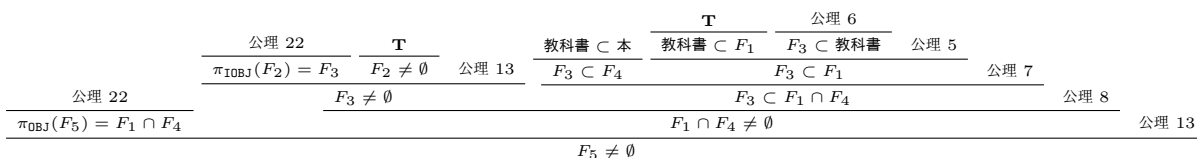


図 4: 外延の抽象表現を用いた証明の例

行うアルゴリズムが知られている [Zhang 96]。さらに、言明は原子文で一階の推論規則は抽象表現の代数的性質を記述する表 2 の公理だけである。大量の原子文に対して限られた推論規則を用いて前向き推論を行う設定はエキスパートシステム [Jackson 98] として古くから研究されており、様々な高速化手法が提案されている。例えば、適切なハッシングにより推論規則の前提を満たす原子文を瞬時に探索することができ [Hanson 90]、効率的な推論エンジンを構築することができる。

図 3 の例では、以下のような抽象表現が構成され、

太郎が読むもの：

$$F_1 = \pi_{IOBJ}(\text{読む} \cap (\text{太郎}_{\text{SBJ}} \times W_{\text{OBJ}}))$$

竹取物語は教科書にある：

$$F_2 = \text{ある} \cap (\text{竹取物語}_{\text{SBJ}} \times \text{教科書}_{\text{IOBJ}})$$

竹取物語のある教科書：

$$F_3 = \text{教科書} \cap \pi_{IOBJ}(\text{ある} \cap (\text{竹取物語}_{\text{SBJ}} \times W_{\text{IOBJ}}))$$

竹取物語のある本：

$$F_4 = \text{本} \cap \pi_{IOBJ}(\text{ある} \cap (\text{竹取物語}_{\text{SBJ}} \times W_{\text{IOBJ}}))$$

太郎は竹取物語のある本を読む：

$$F_5 = \text{読む} \cap (\text{太郎}_{\text{SBJ}} \times F_{4,\text{OBJ}})$$

図 3 左を **T**、右を **H** とすると、**T** の言明 (教科書 $\subset F_1, F_2 \neq \emptyset$)、言語知識 (教科書 \subset 本)、公理から、**H** の表す言明 $F_5 \neq \emptyset$ が証明される (図 4)。

4. 実験

我々は外延の抽象表現に基づく推論エンジンを実装し、英語の含意関係認識データセット PASCAL RTE [Bentivogli 09] および日本語のセンター試験歴史問題について評価実験を行った。

4.1 外延の抽象表現の効率性

PASCAL RTE は、テキスト間含意関係認識の評価タスクで提供されているデータであり、新聞などのテキストから「根拠テキスト-仮説文」ペアを作成したものである。このような実世界のテキストを使った評価データでは、言い換えパタンのバリエーションが非常に大きいため、辞書などから抽出した言語知識だけを使って厳密な証明を試みようとするほとんどの例は証明できない。そのため、我々は動的知識生成のコンポーネントを導入した [田 14]。これは、証明したい仮説文と根拠になるテキスト文との間に、単語類似度に基づくアラインメントをとり、そのアラインメントから論理知識を生成する手法である。この動的知識によって証明できる例は大きく増えたが [田 14]、知識の増加で証明探索のコストも増加した。

実際の証明コストを測るために、我々は RTE5 の開発データから、動的知識を使って証明できる 110 個の正例を抽出し、外延の抽象表現に基づく推論エンジンが証明するのに要した時間を測った (図 5)。図 5 において縦軸は時間 (秒) であり横

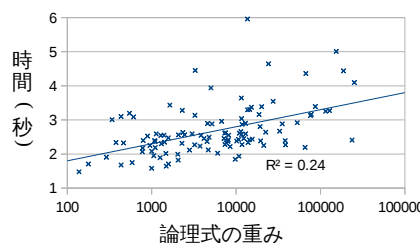


図 5: 論理式の重み (対数標度) に対して我々の推論エンジンが要する証明時間 (秒)

	3 秒以内	5 分以内	5 分以内 (必要知識のみ)
証明できる	8	16	82
変数の数制限	5	24	3
証明探索失敗	0	1	3
メモリ制限	0	2	0
時間制限	86	57	13

表 3: Prover9 の終了状態の割合 (%)

軸は仮説と根拠の全ての論理式の重みである。ただし、論理式の重みは、それを FOL 式に変換した時の全ての述語の重み付け和であり、 n 項述語は重み n とする。我々が使ったマシンは 2.27GHz Xeon CPU で、推論エンジンはシングル・スレッドである。図が示す通り、全ての例は 6 秒以内で証明されることが示された。また、証明時間は論理式の重みに対して対数的なスケール・アップを示している。

一方、全ての論理式を FOL 式に変換し、FOL 自動証明器 Prover9*1 を使った証明を試みた。その結果を表 3 に示す。3 秒以内の時間制限を設けると、証明できたのはわずか 8% である (「3 秒以内」の列)。時間制限を 5 分間に伸ばしても、16% しか証明することができなかった (「5 分以内」の列)。これは、通常の FOL 自動証明器では多くの動的知識 (通常数百個) を含むテキスト推論を扱うのは困難であることを示している。そこで、我々の推論エンジンを用いて証明に実際必要な知識だけを選択し (通常 2、3 個)、Prover9 で証明を試みた。すると 5 分間以内に証明できた例の割合は 82% までに上がり (「5 分以内 (必要知識のみ)」の列)、大量の動的知識は証明コストを大幅に増やすことが示された。一方、約 20% の例は依然証明できないことから、FOL を用いたテキスト推論の困難さを示唆している。

4.2 センター試験と代ゼミ模試タスク

我々のテキスト推論システムをセンター試験および代ゼミ模試タスクの歴史科目に応用し、その有効性を評価した。データセットは、2000 年以降の奇数年度のセンター試験世界史 B (本・追試、01・05・09 年度を開発セット、03・07・11 年度をテストセット) の問題と代ゼミ模試の問題から作成されたものである。選択肢四つの選択問題なのでランダム・ベースライ

*1 www.cs.unm.edu/~mccune/prover9/

データセット	正解率 (%)	問題数
センター世界史 (開発)	42.5	174
センター世界史 (テスト)	44.4	142
代ゼミ模試世界史	40	20
代ゼミ模試日本史	32	22

表 4: 歴史試験問題に対する正解率

ンは 25% である。

知識として用いたのは、日本語語彙体系、分類語意表、広辞苑などから抽出した類義語・反義語知識、および歴史教科書から作成した歴史知識である。歴史教科書は、山川出版の「詳説世界史 B」および「詳説日本史 B」を用いた。教科書はトピックごとに分かれていて、各トピックは数パラグラフ (数十文) からなる。各トピックの自然言語文を外延の抽象表現に対する言明に変換し、類義語・反義語知識と合わせて歴史試験の選択肢にある文の証明を試み、それが史実かどうかを判定する。PASCAL RTE データと同様、言い換えパタンのバリエーションは辞書からの言語知識だけではカバーしきれないので、動的に生成された知識を用いた。最終的な判定には、証明できるか否かの他に、選択肢全体が証明できないとしても、どの部分が証明できているか、固有名詞を含む述語項構造が証明できるかといった情報を用いた。

実験の結果を表 4 に示す。ランダム・ベースラインの 25% より高い精度を達成し、提案システムの有効性が示されている。一方、受験生の平均よりは低い正解率であり、さらなる精度向上が必要である。

間違いの主な二つのタイプの例を以下に示す。

4.2.1 DCS 木への変換

我々のシステムは下記の文を間違っして正解として選んだ (正しくはダヴィデではなくモーセ)。

ヘブライ人は、ダヴィデに率いられてエジプトを脱出した。

ゼロ照応による述語項構造 (ヘブライ人が率いられて) が Syncha で認識されなかったのも一因だが、係り受け木から DCS 木への変換ルールはイベント間の関係 (率いられて-脱出した) を正しく捉えられず、独立した二つのイベントとして認識し、それぞれが教科書から証明できる (ヘブライ人は確かにエジプトを脱出し、ダヴィデもヘブライ人の王なのでゼロ照応が正しく認識できれば「ヘブライ人がダヴィデに率えられる」ことも証明できる) ため、誤った答えを導出してしまふ。

4.2.2 固有名詞認識

固有名詞認識は史実判定において大きな力を発揮することが知られている [Kanayama 12]。今回我々のシステムも、固有名詞を含む述語項構造が証明できなかった場合必ず間違いだとする規則を採用した。しかし、この規則は固有名詞認識の間違いで多くの誤答を引き起こす。

ササン朝のホスロー 1 世は、突厥と同盟を結んでフン族を滅ぼした。

正しくは「フン族」ではなく「エフタル」であるが、「フン族」は固有名詞として認識されなかった。実際、センター試験世界史の開発データにおいて固有名詞認識コンポーネントを試した場合、「教科書のどのトピックにおいても選択肢に含まれる全ての固有名詞が共起しない時、間違っした文とする」規則だけで選択肢を一つに絞れた問題は 43 問 (25%) もあったが、その絞れた選択肢が正解である割合は 39% で全体の正解率を下回った。[Kanayama 12] では人手で固有名詞を与えていたが、固有名詞の自動認識の精度向上は直近の大きな課題である。

5. おわりに

我々は自然言語の推論に適した意味表現を提案し、それに基づいた推論エンジンを実装した。PASCAL RTE データセットおよびセンター試験歴史科目問題に対する実験によって、我々の推論エンジンは効率的な論理推論と実世界の自然言語文の正誤判定ができることが示された。これから各コンポーネントのさらなる精度向上が必要であるが、高度なオントロジー知識との組み合わせによる効果も期待できる。

参考文献

- [Beltagy 13] Beltagy, I., Chau, C., Boleda, G., Garrette, D., Erk, K., and Mooney, R.: Montague Meets Markov: Deep Semantics with Probabilistic Logical Form, in **SEM* (2013)
- [Bentivogli 09] Bentivogli, L., Dagan, I., Dang, H. T., Giampiccolo, D., and Magnini, B.: The Fifth PASCAL Recognizing Textual Entailment Challenge, in *TAC* (2009)
- [Bos 06] Bos, J. and Markert, K.: When logical inference helps determining textual entailment (and when it doesn't), in *2nd RTE Workshop* (2006)
- [Hanson 90] Hanson, E. N., Chaabouni, M., Kim, C.-H., and Wang, Y.-W.: A Predicate Matching Algorithm for Database Rule Systems, in *ACM-SIGMOD* (1990)
- [Jackson 98] Jackson, P.: *Introduction to Expert Systems*, Addison-Wesley Longman, Boston (1998)
- [Kanayama 12] Kanayama, H., Miyao, Y., and Prager, J.: Answering Yes/No Questions via Question Inversion, in *COLING* (2012)
- [Liang 11] Liang, P., Jordan, M. I., and Klein, D.: Learning Dependency-Based Compositional Semantics, in *ACL* (2011)
- [MacCartney 08] MacCartney, B. and Manning, C.: Modeling Semantic Containment and Exclusion in Natural Language Inference, in *Coling* (2008)
- [Moldovan 03] Moldovan, D., Clark, C., Harabagiu, S., and Maiorano, S.: COGEX: A Logic Prover for Question Answering, in *NAACL* (2003)
- [Raina 05] Raina, R., Ng, A., and Manning, C.: Robust textual inference via learning and abductive reasoning, in *AAAI* (2005)
- [Schoenmackers 08] Schoenmackers, S., Etzioni, O., and Weld, D.: Scaling Textual Inference to the Web, in *EMNLP* (2008)
- [Zhang 96] Zhang, H. and Stickel, M. E.: An Efficient Algorithm for Unit Propagation, in *AI-MATH* (1996)
- [川添 12] 川添 愛, 宮尾 祐介, 松崎 拓也, 横野 光, 新井 紀子: 計算機による大学入試問題への解答に向けた世界史オントロジーの設計, 人工知能学会 (2012)
- [田 14] 田 然, 宮尾 祐介: 外延の抽象表現を用いた論理推論, 言語処理学会 (2014)