数学証明における帰結関係を表す接続表現の予測

太田 宇宙　松崎 拓也　藤原 誠
東京理科大学 理学部第一部 応用数学科
1420022@ed.tus.ac.jp {matuzaki,makotofujiwara}@rs.tus.ac.jp

概要

この論文は数学証明において帰結関係を表す hence / therefore / thus の間に使い分けの傾向があるか否かを調べるために、BERT を用いた単語選択の実験を行ったものである。具体的には、arXiv に掲載されている数学論文から証明テキストを抽出し、hence / therefore / thus をすべてマスクした上で BERT を用いて 3 択の穴埋め問題を解かせる訓練及び評価を行なった。

結果として hence / thus の間の使い分けは Wikipedia と BookCorpus のみで事前訓練を行った BERT と比べて向上したが、therefore の正答率のみ hence / thus と比べて低いままだった。故に BERT をベースとしたモデルは少なくとも hence / thus 間の使われ方の違いをある程度まで捉えていると考えられる。

1　はじめに

論理的整合性が重視される数学的な論証において接続表現は重要な役割を果たす。特に「したがって〜である」といった順接 (演繹の帰結) を表す表現は証明において最も基礎的なものであり、それを用いずに書くことは困難である。英語でこれに対応する代表的な単語は hence / therefore / thus である。これらの語の使い分けについて、数学の証明の場合に即して Paquette [1] の説明を解釈すると、おおよそ thus は原因と結果、therefore はある結果が得られる理由の指摘、hence は論理的なつながりを示しているということになる。つまり一部の native speaker はある一定の基準で使い分けを行っていることがわかる。本研究では、BERT によって thus / therefore / hence の使い分けがどの程度可能かを測定することで、数学証明においてこれらの間に一定の使い分けの傾向があるのかどうかを調べた。

具体的には、証明テキスト中の hence / therefore / thus を全てマスクした状態で BERT に入力し、穴埋め形式の 3 択問題を解かせる訓練および評価を行った。本研究は接続表現を対象としているため、両方向の文脈を参照するアーキテクチャを採用している BERT はこのタスクに適していると考えられる。

実験の結果、hence / thus 間の選択に関しては学習前と比べ精度の向上が見られた。そのため、 BERT は少なくとも hence / thus 間の違いをある程度捉えていると考えられる。一方で、therefore をマスクした位置に対し hence と予想するケースが多く見られた。

さらに、hence / therefore / thus の使い分けに関して英語を母語としない数学者による予測実験も行った。BERT による予測の正答率が 46.2% であったのに対し、人間による予測の正答率は 46.1% と、ほぼ同等の精度であった。また、hence / therefore / thus 以外の単語を対象に行った実験では、用法に違いがある since / because に関して BERT による since の再現率 (recall) が 18.04% という非常に低い値となった。この結果から、therefore と同様に since も使い分けの傾向を正確に捉えられていないと判断した。

2　Hence / Therefore / Thus の違い

英語における接続表現 hence / therefore / thus 間にはそれぞれ固有の意味が存在し、置き換え可能である文とそうではない文が存在する。Paquette [1] によると、3 語のうち thus のみが持つ意味は in (this / that) way と with (this / that) (fact / situation / result) であり、ある結果を導くために用いられることが多く、この結果が自然に、かつ必然的に生じるという趣旨を示す。therefore は for (this / that) reason の意味合いが最も強く、はっきりとある結果が出る「理由」を指摘する役割を持つ。hence は as an (inference / deduction) の意味が強く、論理的な脈絡を意味することに hence の特徴があるとしている。数学証明における帰結関係はすべて「必然的」で「はっきりとある結果が出る」「論理的」な関係なので、これらの接続表現は数学証明においてはより同義に近いと言えるだろう。

Oxford English Dictionary [2] では、hence / therefore / thus の意味は以下のように説明されている。

hence As a result of this / consequently / From this source or origin. Now rare. / By inference from these premises or data / for this reason / as a conclusion.

therefore In consequence of that / that being so / consequently.

thus In this way, like this. / In accordance with this / accordingly, and so / consequently / therefore.

3 つ全てに共通する語釈 (consequently) があり、thus の語釈として therefore が示されるなど、論理的帰結を表す点では同義と言えそうだが、同時に理由-推論-帰結の構造の異なる側面に注目しているというニュアンスの差が感じられる。

一方，講義録 “Mathematical Writing” [3] には著者の一人である Knuth によると思しき記述がある:

> There is a definite rhythm in sentences. Read what you have written, and change the wording if it does not flow smoothly. (中略) There are many ways to say “therefore”, but often only one has the correct rhythm.

ここで言われている「リズム」が純粋に音韻的なものを指すのか，あるいは意味合い・ニュアンスの違いも含めた「流れ」があるとの趣旨なのかは明確でないが，いずれにせよ帰結を示す種々の表現のうち，ただ一つだけが適切であるような文脈が往々にして存在すると主張されている．

3 実験設定

プレプリントサーバ arXiv[1)] で公開されている数学分野の論文のうち、TEX ソースを伴って 2023 年 5 月までに公開されたもの全てを使用した。TEX ファイルから直接テキストを抽出するのは困難であったため、一度 XML ファイルに変換してからテキストの抽出を行った。具体的な手順を次に記す。

1. TEX 形式で書かれたファイルを LATEXML[2)] で XML ファイルに変換する。CPU 時間が 300 秒を超えた場合、または変換時に何らかのエラーが発生した場合は変換失敗として終了する。
2. XML 形式のファイルの整形を行う。
   TEX の ref コマンドおよび cite コマンドに相当する部分を [1] と置き換える。
   数式モードに相当する部分 (Math タグ) で、タグの子要素の数が 2 (A , 1 など) のときはそのまま、それ以外の場合は [formula] と置き換える。
3. 整形したデータからタグが TEX の proof 環境に相当する部分を探し、テキストを抽出する。

最終的に 1 つ以上の proof 環境が得られた論文の数は 274,886 個、proof 環境数は 1,510,391 個であった。

4 予測手法

実験では BERT を用いた 4 つの予測手法を比較した。以下でこれらについて説明する。

4.1 予測モデル

すべてのモデルの入力は最大 510 トークンとする。これを超える長さを持つ証明はスライディングウィンドウの方式で入力を行う。すなわち、幅 510 トークンの窓を 255 トークンずつずらしながら全体を入力し、入力中の hence / therefore / thus は全てマスクトークンで置き換える。

4.1.1 事前訓練のみの BERT による MLM

Devlin らによる Wikipedia および BookCorpus で事前訓練を行った BERT (transformers ライブラリの bert-base-uncased) を使用し、hence / therefore / thus のうち MLM のスコアが最も高いものを選ぶ。

4.1.2 BERT による 3 値分類

マスクした各トークンが hence / therefore / thus のいずれであったかを 3 値分類の形で予測する。具体的には、マスクしたトークンに対する BERT の出力ベクトルにパラメータ行列を掛けて 3 つの選択肢に対するスコアを得ることで分類を行う。

3 値分類の訓練の際の初期値を §4.1.1 のモデルとした場合と、§4.1.1 のモデルに対し、数学証明テキスト上での Masked Language Model(MLM) の訓練をさらに行ったものを比較した。以下、前者を追加訓練なしの BERT による 3 値分類と呼び、後者を追加訓練ありの BERT による 3 値分類と呼ぶ。MLM の訓練のエポック数は 3 であり、上記 a. と同じ訓練データを使用している。マスクする確率は 15% にしている。

1) https://arxiv.org/
2) https://math.nist.gov/~BMiller/LaTeXML/

表 1　hence / therefore / thus の選択精度

数学証明での MLM 訓練	モデル	精度 (全体)	再現率		
			hence	therefore	thus
なし	BERT による MLM	34.86%	18.92%	61.95%	30.94%
なし	BERT による 3 値分類	48.38%	61.89%	35.81%	43.69%
あり	BERT による 3 値分類	48.89%	55.95%	34.14%	52.53%
なし	BERT+LSTM による 3 値分類	47.62%	60.46%	36.69%	42.4%

表 2　BERT による 3 値分類の混乱行列

予測 \ 正解	hence	therefore	thus
hence	74495	31912	44615
therefore	18475	31421	21001
thus	27392	24403	50917

表 3　BERT+ 証明 MLM による 3 値分類の混乱行列

予測 \ 正解	hence	therefore	thus
hence	67345	26442	36554
therefore	17504	29956	18811
thus	36240	31465	61209

表 4　BERT+LSTM による 3 値分類の混乱行列

予測 \ 正解	hence	therefore	thus
hence	72766	31814	44175
therefore	19074	32194	22933
thus	28097	23722	49411

表 5　人間による予測の混乱行列

予測＼正解	hence	therefore	thus
hence	17	12	15
therefore	8	20	19
thus	10	13	16

表 6　機械による予測の混乱行列

予測＼正解	hence	therefore	thus
hence	5	5	4
therefore	1	3	2
thus	1	1	4

表 7　年代ごとの分布

年代	hence	therefore	thus	正答率
90 年代	41.13	22.4	36.47	48.18
00 年代	38.75	24.81	36.45	52.25
10 年代	39.07	26.22	34.71	51.02
20 年代	38.53	26.67	34.81	51.09

4.1.3　BERT+LSTM による 3 値分類

マスクしたトークンに対する BERT の出力ベクトルに一つ前のマスク位置に対する出力の埋め込みを結合し、LSTM に入力して 3 つの選択肢に対するスコアを得る。訓練の際の初期値として §4.1.1 のモデルを用いた。

4.2　訓練設定

訓練は、最大エポック数 5、バッチサイズ 16、学習率 5×10^{-6}、損失関数を交差エントロピーとした。また、全データを 8:1:1 に分け、訓練データ、検証データ、テストデータとした。各エポック後に検証データについて損失を計算し、最も小さくなった時点のパラメータを用いてテストを行った。

5　実験結果

前節で説明した 4 つの方法による hence / therefore / thus の選択の精度を表 1 に示す。また表 2,3,4 に 3 値分類の形式で訓練した 3 つのモデルによる予測結果の混乱行列を示す。混乱行列の行は予測した接続表現、列は正解の接続表現を示している。全データ中の hence / therefore / thus の総数は順に 120,362 、87,736、116,533 であった。hence を常に選んだ場合、全体の正答率は 37.08% となる。上三つの方法は正答率が 48% 程度を記録し、therefore の再現率が hence / thus と比べて低い。対照的に追加訓練なしの MLM は正答率が低いが、therefore の再現率が hence / thus と比べて高いことが表から読み取れる。

6　人間による予測精度

前節で用いたテストデータのうち論理学分野の論文から抽出した証明 10 個から 26 問の穴埋め問題を作り、日本語ネイティブの数学研究者 5 名に証明テキスト以外は参照せずに解いてもらった。1 問 1 点であり、26 点が満点となる。

結果は平均値 10.6 点、中央値 11 点、平均正答率は 46% となった。hence / therefore / thus に対する再現率は順に 48%, 44%, 32% であった。一方、5 節で最も精度の高かった、追加訓練ありの BERT による 3 値分類のモデルに同じ問題を解かせたところ、正答率は 46.15% となった。hence / therefore / thus に対する再現率は順に 71%, 33%, 40% である。図 5 および図 6 に、人間および BERT を用いた予想結果の混乱行列を示す。平均正答率は人間と BERT で近いが、誤りの傾向は必ずしも同様でないことがわかる。

7　分析

数学証明における接続表現 hence / therefore / thus の選択がもしも純粋に文体 (スタイル) 上のものであれば、分野や年代によって 3 つの語の相対頻度およびそれぞれが選択される文脈タイプは大きく異なる可能性がある。これを検証するために年代、分野によって接続表現の予測精度および相対頻度が変化するかを確かめた。また、接続表現の予測に関する実験結果を比較するため、他の単語の組み合わせを対象にして予測精度を調べた。

表 8　系列の影響

		次にくる接続表現		
		hence	therefore	thus
基準の接続表現	hence	47.03%	24.01%	28.96%
	therefore	32.00%	39.31%	28.68%
	thus	29.56%	21.97%	48.47%

表 9　他単語間の精度

選択肢	精度	ベースライン (最多の単語)
since/because/as	85.25%	47.13%(as)
since/because	89.52%	87.46%(because)
cf./see	99.80%	92.64%(see)
a/the	96.36%	72.94%(the)
hence/a/to	99.64%	39.61%(a)

表 10　since / because / as の混乱行列

予測＼正解	since	because	as
since	4939	857	246
because	18741	171321	29480
as	1441	5932	151425

表 11　since / because の混乱行列

予測＼正解	since	because
since	4532	711
because	20589	177399

7.1　hence / therefore / thus の分布

年代ごとの分布は表 7 の通りとなった。表中の正答率は、追加訓練ありの BERT による 3 値分類のモデルによる予測結果についてのものであり、テストデータを年代別に分けた際の正答率である。年代別では大きな違いはないと考えられる。分野別では全体の正答率の範囲が 46% ~ 56% 程度となった。正答率に少々揺れが出ているが、大きな違いはないと考えられる。詳細は付録に掲載する。

7.2　系列の影響

テキストデータ全体から hence / therefore / thus のみを抽出し、同じ証明内に出現する接続表現の組み合わせに傾向があるかどうかを調べた。表 8 の行は先行する接続表現、列はその次に出現した接続表現の割合を示している。例えば 2 行 1 列は therefore の次に hence が現れる相対頻度を示している。

表 8 によれば、hence / thus に関しては同じ接続表現を比較的に使う傾向にある。

7.3　他の単語間の選択との比較

接続表現の選択の難しさを他の単語の組の間の選択と比較した。追加訓練なしの BERT による 3 値分類を他の単語の組合わせについて訓練したものを用いた。全ての実験結果を表 9 に示す。ベースラインは、単語の組のうち最多の単語の比率を示す。

7.3.1　since / because(/ as)

正答率は 85.25% となった。最も多い as を常に選んだ場合、正答率は 47.13% となる。予測結果の混乱行列を表 10 に示す。

選択肢から as を除いた、since / because についても実験も行った。because を常に選んだ場合、正答率は 87.64% となる。これに対し実際の正答率は 89.52% となった。予測結果の混乱行列を表 11 に示す。

7.3.2　hence / therefore / thus との比較

選択肢とする語のうち cf. / see と a / the に注目すると、どちらも予測精度が高い。前者に関しては cf. は「比較のために参照する」、see は「(単に) 参照する」という意味の違いがあるが、「参照せよ」の意味ではほぼ同じ位置で使われる単語である。後者は可算名詞や非可算名詞など、複数の条件がある上で使い分けをする必要があるが、ほぼ同じ位置で使われる単語である。これらの違いは BERT に基づく統計モデルで比較的正確に捉えられると考えられる。

一方で、since / because(/ as) については since の再現率が非常に低いという、他とは違う結果が得られた。3 単語の違いに関して興野 [4] は because は新情報として明確な理由を述べる、since はすでに分かっている理由を述べる、as は補足的理由を述べる、と説明している。いずれも副詞で、同じ位置に出現可能な hence / therefore / thus とは違い、主に since は文頭、because は文末に用いられる。このように since / because に関してはより多くの判断材料があるにも関わらず、精度が低い。この結果は、since / because の使い分けは傾向が明確でないことが示唆されている。therefore と hence / thus の使い分けに関しても同様のことが言える。

8　おわりに

これまでの結果から、therefore と hence / thus 間の違いを BERT に基づく統計モデルは正確に捉えられていないと言える。そのため、therefore と他の 2 語の使い分けは、より傾向が明確でないことが示唆される。この研究では置き換え可能な文が混在している状態で実験を行った。これが therefore と hence / thus 間の違いを正確に捉えられていないという結果に繋がったのか、あるいはさらに別の要因があったのかを突き止めるのは今後の課題である。

参考文献

[1] Glenn Paquette. 科学論文の英語用法百科: English composition for scholarly works. よく誤用される単語と表現. 第 1 編. 京都大学学術出版会, 2004.

[2] オックスフォード英語辞典. Oxford english dictionary, 2023. oed.com.

[3] Donald E. Knuth, Tracy Larrabee, and Paul M. Roberts. **Mathematical Writing**, Vol. 14 of **MAA notes**. Mathematical Association of America, 1989.

[4] 興野登.【英語論文の書き方】第 24 回　because, since, as など 理由を表す表現, 2016. https://www.worldts.com/english-writing/eigo-ronbun24/index.html.

A　分野ごとの hence / therefore / thus の分布

7.1 節の hence / therefore / thus の分布において、年代別の他に分野別でも集計を行っていた。分野ごとの分布は付録の表 12 の通りになった。複数分野に及ぶ論文はそれぞれの分野に加算している。また、分野名は arXiv の表記に従う。対象となる分野の論文の数が小さいものも含まれている影響で、正答率に少々揺れが出ているが、分野別でも正答率に大きな違いはないと考えられる。

表 12　分野ごとの分布

分野名	hence	therefore	thus	正答率
math.CA	35.45	28.16	36.4	47.55
math.FA	40.33	26.77	32.91	49.29
math.NT	38.88	26.13	34.99	52.26
math.OA	40.64	23.94	35.43	51.86
math.CO	35.84	24.36	39.8	54.39
math.AG	42.64	24.24	33.11	54.44
math.DG	38.54	27.19	34.27	50.17
math.AC	40.85	24.79	34.37	52.11
math.AP	36.07	28.14	35.79	48.42
math.QA	40.02	24.38	35.6	52.67
math.RT	41.68	24.17	34.15	52.38
math.GR	41.75	24.19	34.06	55.26
math.MG	39.34	25.5	35.16	49.71
math.AT	38.75	27.04	34.21	51.17
math.CT	39.44	23.51	37.04	54.00
math.PR	35.56	28.57	35.87	48.08
math-ph	36.15	27.08	36.77	50.17
math.MP	36.15	27.08	36.77	50.17
math.ST	36.34	29.28	34.37	44.81
math.RA	41.31	23.83	34.85	51.53
math.OC	38.17	28.1	33.73	46.56
math.DS	38.89	26.37	34.74	50.03
math.GT	42.54	25.84	31.62	54.24
math.CV	38.96	26.86	34.18	51.13
math.KT	43.71	24.48	31.81	52.30
math.GM	40.25	28.54	31.2	51.30
math.SG	41.64	26.42	31.93	53.02
math.LO	41.52	24.14	34.34	56.28
math.SP	37.41	26.29	36.3	49.25
math.IT	39.29	27.05	33.66	48.32
math.GN	45.03	25.75	29.22	56.01
math.NA	35.49	28.1	36.41	46.30
math.HO	30.5	30.05	39.45	50.58