

日本語数量表現のためのアノテーション体系と推論データセットの構築

小谷野 華那 (指導教員: 戸次 大介)

1 はじめに

自然言語理解の基礎をなすタスクの一つとして、自然言語推論がある [2, 1]. 自然言語推論は、前提文が真であるとき、仮説文が真ならば含意 (ENTAILMENT), 偽ならば矛盾 (CONTRADICTION), どちらともいえないならば中立 (NEUTRAL) であることを判定するタスクである. 自然言語推論では一般に意味論的推論の判定が想定されているが, 自然言語処理分野では近年, 意味論的推論だけでなく, 語用論的推論も研究の対象となっている [5]. この2種類の推論は, 言語学の文献で議論されてきた含意 (entailment)¹と推意 (implicature) に対応する [6, 4, 7]. 例として, 以下のような数量表現を含む前提文 P と仮説文 H のペアについて考えよう.

- (1) P : 男性が道端に4人座っていた.
 H : 男性が道端に5人座っていた.

この P と H に現れる数量表現には, 「少なくとも4人 (5人) 座っていた」と「ちょうど4人 (5人) 座っていた」という2種類の解釈がありうる. 前者を意味論的解釈, 後者を語用論的解釈と呼ぶ. P から H への推論は, 意味論的解釈のもとでは中立 (NEUTRAL) であるのに対し, 語用論的解釈のもとでは矛盾 (CONTRADICTION) となる. このように数量表現を含む前提文と仮説文のペアが与えられたとき, 数量表現の解釈の仕方によって判定が異なることがあるため, この2通りの解釈を区別して考える必要がある. 以降では, 意味論的解釈に基づく判定ラベルを「含意ラベル」, 語用論的解釈に基づく判定ラベルを「推意ラベル」と呼ぶ.

日本語の数量表現には様々な助数辞が存在し, 出現形式も多様である. (1) の P の助数辞は「名」に変更することができ, 「4人の男性が道端に座っていた」と語順を入れ替えることも可能である. こうした数量表現の出現形式や助数辞の多様性は, 日本語の重要な特徴の一つであり, 言語学での記述的・理論的研究が近年進んでいるが, これらの特徴に着目したコーパスの構築や, 言語学的な知見をふまえて日本語の数量表現の理解を問うようなデータセットの構築は, 管見の限り行われていない.

本研究では, 助数辞の種類, 数量表現の出現形式, 用法をアノテーションした日本語の数量表現アノテーションコーパスを構築する. さらに, 作成した日本語数量表現アノテーションコーパスをもとに, 含意ラベルと推意ラベルを付与した数量表現の推論データセットを構築する. 本稿では, 数量表現コーパスおよび推論データセットの構築と, 構築した2つのコーパスを用いて, 現在の標準的な事前学習済み言語モデルが, 事前学習時に数量表現の振る舞いを正しく学習できているか, 数量表現の理解を必要とする推論をどの程度扱えるかを調査する実験を行う.

¹広い意味での含意ではなく, 推意 (implicature) と対比した論理的含意 (entailment) を指す.

2 数量表現コーパスの構築

本研究では, NPCMJ [9] 内の126文に含まれる287件の数量表現について, 言語学の素養のある大学院生1名がアノテーションを行った. 複雑な文を扱った挑戦的な推論データセットを構築することを最終目的として, 数量表現を2つ以上含む文をアノテーションの対象として抽出した. また, 否定表現と条件節を含む文については数量表現を1つ以上含む文を抽出した. 文中に現れる数量表現に<num>タグを付与し, 助数辞の分類, 出現形式, 用法についてアノテーションを行った. 助数辞, 出現形式, 用法のタイプの例と件数を表1に示す.

表1: (a) 助数辞, (b) 出現形式, (c) 用法の分類の例とアノテーション件数

	タイプ	例	件数
a	分類辞	人, 頭, 冊, 枚	62
	単位形成辞	瓶, 箱, 袋, バック, 切れ	11
	計量辞	リットル, 円, バイト	106
	順序数辞	月, 日, 番, 位	108
b	Q / NC型	3人の学生が来た	34
	N / QC型	学生の3人が来た	13
	NCQ型	学生が3人来た	33
	NQC型	学生3人が来た	10
	テ格型	学生が3人で来た	9
	述部型	来た学生は3人だ.	26
	QV型	東京に3回行った	86
	NvCQ型	渡米したことは2回ある	1
	Nの脱落	3人はお金を払った	34
	QQC型	1時間500円かかる	8
	(Q)	(1990年)	15
イディオムの	1人暮らし, 8人兄弟	18	
c	QがNのカテゴリ情報を表すもの	3人の学生	74
	QがNを構成する要素の全体数を表すもの	家族3人	4
	QがNを構成する要素の一部を表すもの	集団の1人	7
	QがNの属性や特徴を表すもの	50歳の男性	74
	Vが行われた回数を表すQ	2回来る	1
	Vが行われた期間を表すQ	3日滞在する	29
	Vが行われた時間を表すQ	9時に来る	57
	Vの特徴を表すQ	2%増加する	11
	Nvを修飾するQ	渡航歴は2回	4
	Qを修飾するQ	1時間500円	8
	イディオムの用法	1人暮らし	18

表2: 推論データセットの統計情報

	ENTAILMENT	CONTRADICTION	NEUTRAL
含意ラベル	526	634	356
推意ラベル	527	793	196

3 推論データセットの構築

数量表現コーパスを用いて, 数量に関する推論データセットを構築した. 数量表現コーパスに含まれる文を前提文 P とした. 仮説文は, <num>タグが付与されている前提文の数量表現について, 意味を変えない最小の節を取り出し, <num>タグが付与されている数詞の変更と数量詞修飾を付与することで作成する. 数詞の変更は, 数量表現に応じて増やす数, 減らす数を変更して, 非文にならない範囲で語順の変更も行っている.

推論データセットの各ペアに対して含意ラベルと推意ラベルを付与し, 推論データセットを作成した. 含意ラ

表 3: 推論データセットに含まれる文ペアの例

前提文 P と仮説文 (H_- , H_+)	正解ラベル			
	(P, H_-)		(P, H_+)	
	含意	推意	含意	推意
P : 前回 1997 年の税率アップ時を参考にすれば, 昨年 12 月~3 月の駆け込み需要で前年比 1 % の売り上げ増が見込まれる半面, ことし 4 月以降は 4~5 % 程度の落ち込みが予想される. H_- : 前回の税率アップは 1996 年より後だった. H_+ : 前回の税率アップは 1998 年より後だった.	ENTAILMENT	ENTAILMENT	CONTRADICTION	CONTRADICTION
P : 勿論, 私ひとりでは四升呑みほしたわけではない. H_- : 勿論, 私ひとりでは三升以上呑みほしたわけではない. H_+ : 勿論, 私ひとりでは五升以上呑みほしたわけではない.	NEUTRAL	NEUTRAL	ENTAILMENT	ENTAILMENT

表 4: 東北大 BERT を用いた評価実験の結果 (正答率)

学習データ	含意ラベル				推意ラベル			
	all	ENTAILMENT	CONTRADICTION	NEUTRAL	all	ENTAILMENT	CONTRADICTION	NEUTRAL
JSICK	32.45%	69.39%	10.57%	16.85%	31.53%	69.45%	10.21%	15.82%
JSNLI	40.17%	65.40%	37.70%	7.30%	42.68%	65.09%	36.32%	8.16%

ベルと推意ラベルはそれぞれ含意 (ENTAILMENT), 矛盾 (CONTRADICTION), 中立 (NEUTRAL) の 3 値である.

本研究で作成した推論データセットは, 1,516 件の前提文と仮説文のペアを含む. 推論データセットの統計情報を表 2 に, データセットに含まれる文ペアの例を表 3 に示す.

4 事前学習済み言語モデルの評価実験

助数辞予測 数量表現コーパスから助数辞をマスクしたテストセットを構築し, 事前学習済み言語モデルが助数辞を正しく予測できるかを調査する実験を行う. テストセットは, 助数辞をマスクした文に対して, 正解の助数辞と誤りの助数辞を準備し, 言語モデルが誤りの助数辞よりも正解の助数辞を高い確率で予測するかを調査する. 実験には, 東北大 BERT² [3] と早大 RoBERTa³ [8] を用いる. 実験結果を表 5 に示す. 助数辞ごとの正答率では, 東北大 BERT と早大 RoBERTa のどちらも単位形成辞の正答率がやや低いものの, 全体としては東北大 BERT は 89.00%, 早稲田 RoBERTa は 94.04% と高い正答率である.

推論 現在の標準的な事前学習済み言語モデルが数量表現の理解を必要とする推論をどの程度扱えるかを評価するため, 東北大 BERT の評価実験を実施した. 実験結果を表 4 に示す. 含意ラベル, 推意ラベルともに, JSICK [10], JSNLI [11] どちらで学習した場合も, 全体の正答率は 5 割以下と低くなっている. また, ENTAILMENT は正答率が 6 割以上であるのに対し, CONTRADICTION, NEUTRAL は 4 割以下という結果であり, 東北大 BERT は ENTAILMENT と予測する傾向があると考えられる.

5 おわりに

本研究では, 言語学の理論に基づいた日本語数量表現コーパスと推論データセットを構築し, 現在の標準的な事前学習済み言語モデルが数量表現の振る舞いを正しく学習できているか, 数量表現の理解を必要とする推論を正しく扱うことができるかを調査し, 事前学

表 5: 助数辞ごとの評価実験の結果 (正答率)

タイプ	東北大 BERT	早稲田 RoBERTa
ALL	89.40%	94.04%
分類辞	82.50%	92.50%
単位形成辞	75.00%	75.00%
計量辞	91.49%	93.62%
順序数辞	94.64%	98.21%

習済み言語モデルには, 日本語数量表現の扱いに課題がある可能性を確認した.

謝辞 本研究の一部は, JST CREST JPMJCR20D2, JST さきがけ JPMJPR21C8 の支援を受けたものである.

参考文献

- [1] Bowman, S. R., Angeli, G., Potts, C. and Manning, C. D.: 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 (2015).
- [2] Cooper, R., Crouch, R., Eijck, van J., Fox, C., Genabith, van J., Jaspers, J., Kamp, H., Pinkal, M., Poesio, M., Pulman, S., et al.: FraCaS—A Framework for Computational Semantics, *Deliverable*, Vol. D6, (1994).
- [3] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (2019).
- [4] Horn, L. R.: *A Natural History of Negation*, University of Chicago Press (1989).
- [5] Jeretic, P., Warstadt, A., Bhooshan, S. and Williams, A.: Are Natural Language Inference Models IMPPRESsive? Learning IMPLICature and PRESupposition, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 8690–8705 (2020).
- [6] Levinson, S. C.: *Pragmatics*, Cambridge University Press (1983), 安井稔・奥田夏子訳『英語語用論』研究社出版, 1990 年.
- [7] Levinson, S. C.: *Presumptive Meanings: The Theory of Generalized Conversational Implicature*, MIT press (2000).
- [8] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V.: RoBERTa: A Robustly Optimized BERT Pretraining Approach, *ArXiv*, Vol. abs/1907.11692, (2019).
- [9] NINJAL, : NINJAL Parsed Corpus of Modern Japanese. (Version 1.0) (2016), <https://npcmj.ninjal.ac.jp/>.
- [10] Yanaka, H. and Mineshima, K.: Compositional Evaluation on Japanese Textual Entailment and Similarity, *Transactions of the Association for Computational Linguistics* (2022), to appear.
- [11] 吉越卓見, 河原大輔, 黒橋禎夫 F 機械翻訳を用いた自然言語推論データセットの多言語化, 第 244 回自然言語処理研究会 (2020).

²<https://huggingface.co/cl-tohoku/bert-base-japanese>

³<https://huggingface.co/nlp-waseda/roberta-base-japanese-with-auto-jumanpp>