供述調書に現れる数量表現の推論テストセットの構築

小谷野華那[1]　谷中瞳[2]　峯島宏次[3]　福田浩司[4]　橋爪宏典[5]　戸次大介[1]

1. お茶の水女子大学　2. 東京大学　3. 慶應義塾大学

4. 日本電気株式会社　5. NECソリューションイノベータ株式会社

JED2022　2022年3月18日

Agenda

1. はじめに
2. 関連研究
3. 数量表現の分類
    a. 助数辞の分類
    b. 数量表現の出現位置
    c. 数量表現の用法
4. テストセットの構築
5. 評価
6. おわりに

1. はじめに

自然言語処理の応用

刑事手続文書への応用

数量表現の意味を正しく処理することが重要

供述調書

数量表現が頻出

数量表現の意味を正しく処理する言語処理技術の開発・評価

実テキストの数量表現の理解を問うデータセットが必要

1. はじめに

タバコを一日一箱吸う。

最低でも一箱吸う

一箱は吸うけど
二箱までは吸わない

含意

推意

文脈や状況によって、解釈に揺れがある場合がある

1. はじめに

含意関係認識

前提文が真であるときに仮説文が真(yes)か偽(no)かどちらともいえないか(unknown)を判定するタスク

文ペア

T：タバコを一日一箱吸う。

H：タバコを一日二箱以上吸う。

含意

Tの解釈：

最低でも一箱は吸う

➡ unknown

推意

Tの解釈：

一箱は吸うけど二箱は吸わない

➡ no

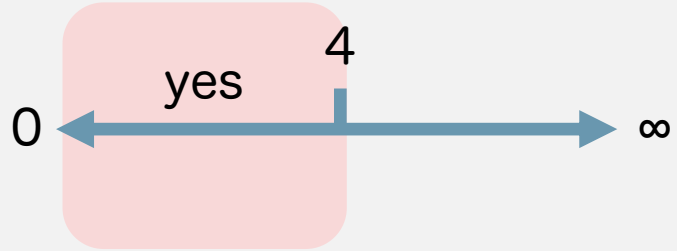
文の解釈の違い・数詞の違い・数量表現の用法によって含意と推意の間で判断が異なる

1. はじめに

否定文・条件文は平叙文の時と判定ラベルが異なる場合がある

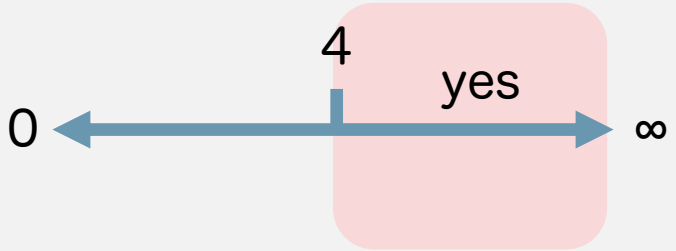
平叙文

T　：４人いる。
H-　：３人いる。
H+　：５人いる。



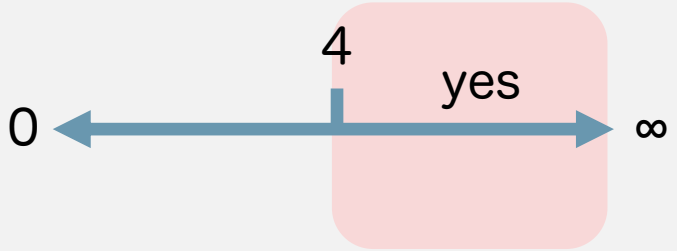
否定文

T　：４人は**いない**。
H-　：３人は**いない**。
H+　：５人は**いない**。



条件文

T　：４人**いれば**できる。
H-　：３人**いれば**できる。
H+　：５人**いれば**できる。



1. はじめに

目標 意味アノテーション付きの数量表現コーパスの構築と、それに基づく推論テストセットの構築

数量表現コーパス

テキスト

- 刑事手続に関連する実テキストのうち数量表現を含む文
- NPCMJ[2]から抽出した数量表現を含む文

アノテーション

- 数量表現の分類や用法

推論データセット

2つの推論ラベルを付与したテストセット

[2] NINJAL. NINJAL Parsed Corpus of Modern Japanese. (Version 1.0). Technical report, National Institute for Japanese Language and Linguistics, 2016. https://npcmj.ninjal.ac.jp/.

2. 関連研究

英語の推論データセット

Naik+, 2018

- 英語の数量表現を含む推論テストセット
- 含意、矛盾、中立の3つのカテゴリ：
  計7596件の文ペア（各2532件）

Jeretic+, 2020

- 英語のscalar implicatureのデータセット
- テンプレートから自動で構築
  →比較的単純な文が多い。

批判

Liu+, 2019

いくつかのヒューリスティクスで
全体の約82%を解くことができる

2. 関連研究

日本語の推論データセット

JSeM[Kawazoe+, 2017]

形式意味論テストセット

JSNLI[吉越+, 2020]

英語SNLI[Bowman+, 2015]の日本語版

JSICK[谷中+, 2021]

英語SICK[Marelli+, 2014]の日本語版

JRTEC[Hayashibe+, 2020]

旅行情報サイトの評価（実テキスト）から
クラウドソーシングで構築

数量表現のデータ

成澤+, 2012

- 日本語の含意関係認識において数量表現が問題になる事例に焦点を当てて分析
- 数量表現の規格化のためのモジュールの実装と評価
- 数量表現が出現する文ペアを7つのカテゴリに分類
- 含意関係を正しく判定するために必要な処理

数量表現の分類・数詞の違いによる
含意と推意の判定ラベルの違いについての
言及はない

3. 数量表現の分類

助数辞の分類

飯田, 2019

- 分類辞
- 単位形成辞
- 計量辞

奥津, 2019

- 序数辞

出現位置

岩田, 2013

- QノNC型
- NノQC型
- NCQ型
- NQC型
- デ格型
- 述部型

新たに追加

- QV型
- NvCQ型
- イディオム的
- Nの脱落
- QtQ型

用法

岩田, 2013

- QがNのカテゴリー情報を表す

新たに追加

- QがNを構成する要素の全体数を表す
- QがNを構成する要素の一部を表す
- QがNの属性や特徴を表す
- Vが行われた回数を表すQ
- Vが行われた期間を表すQ
- Vが行われた時間を表すQ
- Vの特徴を表すQ
- Nvを修飾するQ
- イディオム的用法

3-a. 助数辞の分類

判定　助数辞の後ろに「分」をつける

飯田, 2019

分類辞

- 助数辞単体では使われない
- 複数の人やものに対して使われるものも含まれる

Ex) 人、頭、冊、枚、組

そのままでは何を意味しているかわからない

T：3人の子供がいる。
H：3人分の子供がいる。

単位形成辞

- 容器を表すもの
- より大きなものの部分を指すもの

Ex) 瓶、箱、パック、切れ

Tが正しければHも正しい
その逆は言えない

T：本が3箱ある。
H：本が3箱分ある。

計量辞

- 計量の単位を表すもの

Ex) リットル、円、バイト

Tが正しければHも正しい
Hが正しければTも正しい

T：水が2リットルある。
H：水が2リットル分ある。

3-a. 助数辞の分類

判定　助数辞の後ろに「分」をつける

飯田 2019

分類辞

- 助数辞単体では使われない
- 複数の人やものに対して使われるものも含まれる

Ex) 人、頭、冊、枚、組

単位形成辞

- 容器を表すもの
- より大きなものの部分を指す助数辞

Ex) 瓶、箱、パック、切れ

計量辞

- 計量の単位を表す

Ex) リットル、円、バイト

奥津, 1996

序数辞

- 特定の値を表す助数辞
- 飯田の分類テストを用いて「分」をつけると非文になる

Ex) 月、日、番

4日にイベントがある  4日分にイベントがある

3-a. 助数辞の分類

A：ビルの3階にのぼる。
B：ビルを3階のぼる。

A

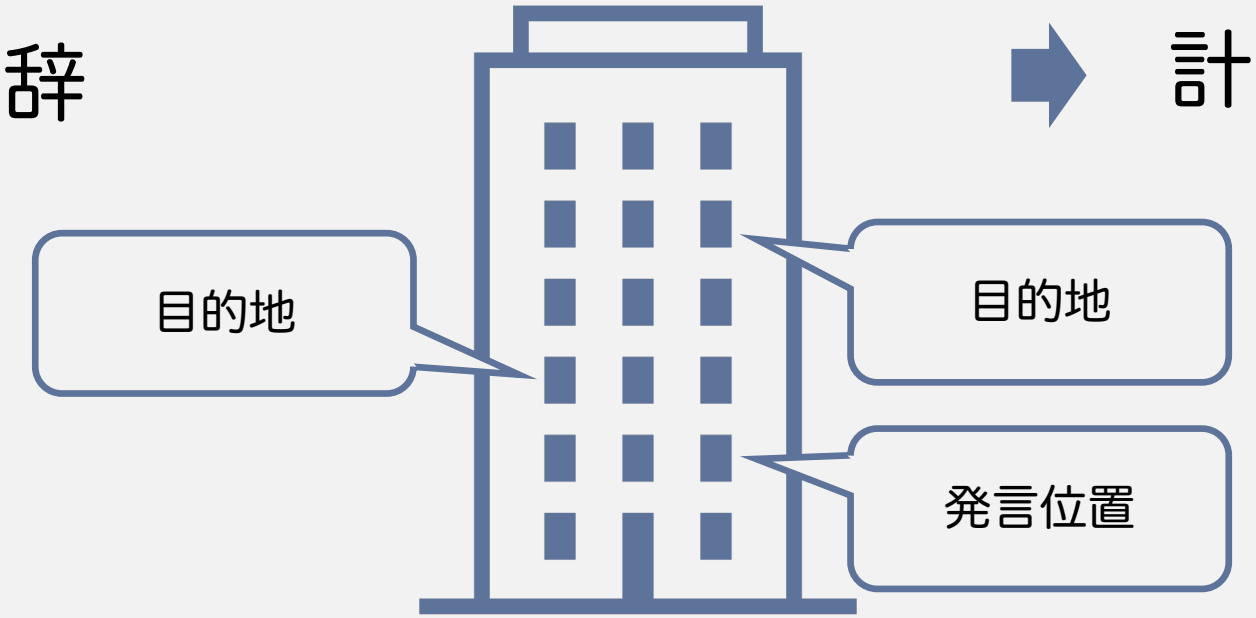
ビルの3階にのぼる。
＝ 建物の3階部分にいく。

➡ 序数辞

B

ビルを3階のぼる。
＝ 発言位置から3フロア上に行く。

➡ 計量辞



3-b. 数量表現の出現位置

岩田 2013

Nについてその数とカテゴリー情報をQが表すもの

N（Noun）	名詞
Q（Case）	数量表現
C（Quantifier）	格助詞

QノNC型	3人の学生が来た。
NノQC型	学生の3人が来た。
NCQ型	学生が3人来た。
NQC型	学生3人が来た。
述部型	来た学生は3人だ。
デ格型	学生が3人で来た。

新たに追加

V（Verb）	動詞
Nv（Verb Stem）	イベント名詞句

QV型	東京に3回行く。
NvCQ型	渡航したことは3回ある。
イディオム的	1人暮らし

3-c. 数量表現の用法

Nに対するQ

QがNのカテゴリーを表すもの	3人の学生
QがNを構成する要素の全体数を表すもの	家族3人
QがNを構成する要素の一部を表すもの	集団の1人
QがNの属性や特徴を表すもの	50歳の男性

Vに対するQ

Vが行われた回数を表すQ	2回来る
Vが行われた期間を表すQ	3日滞在する
Vが行われた時間を表すQ	9時にくる
Vの特徴を表すQ	2%増加する

Nvを修飾するQ　渡航歴は2回

イディオム的用法：1人暮らし

4. データセットの構築

1 フォーカスする数量表現一つにタグ付け

意味アノテーション付き数量表現コーパスの構築

2 仮説文を作成

3 含意ラベルと推意ラベルを付与

推論データセットの構築

4. データセットの構築

1 フォーカスする数量表現一つにタグ付け

2 仮説文を作成

3 含意ラベルと推意ラベルを付与

河川敷では10人の男性がバーベキューをしていた。

・・・では<num>10人</num>の男性が・・・

分類	出現位置	用法
分類辞	QノNC型	QがNのカテゴリーを表すもの

チェック項目

代名詞的用法	数詞1	遊離数量詞	Qが裸
「のうち・中」	「含め」	量化	一定期間
漢数字	negation	conditional	

アノテーションした数量表現	供述調書	431件
	NPCMJ	313件

4. データセットの構築

1 フォーカスする数量表現一つにタグ付け

2 仮説文を作成

3 含意ラベルと推意ラベルを付与

河川敷では10人の男性がバーベキューをしていた。

⬇ 数字をプラスマイナス＋numeral modifier

河川敷では9人以上の男性がバーベキューをしていた。
河川敷では11人以上の男性がバーベキューをしていた。

河川敷では10人くらいの男性がバーベキューをしていた。

⬇ unknownにならないように数詞を動かす

河川敷では2人以上の男性がバーベキューをしていた。
河川敷では20人以上の男性がバーベキューをしていた。

以上系	778件
以下系	583件
ちょうど系	572件

4. データセットの構築

1 フォーカスする数量表現一つにタグ付け

2 仮説文を作成

3 含意ラベルと推意ラベルを付与

T：河川敷では10人の男性がバーベキューをしていた。

H1：河川敷では9人以上の男性がバーベキューをしていた。
H2：河川敷では11人以上の男性がバーベキューをしていた。

	T⇒H1	T⇒H2
含意ラベル	yes	unknown
推意ラベル	yes	no

	yes	no	unknown
含意ラベル	653件	660件	620件
推意ラベル	656件	1092件	185件

5. 評価

モデル　日本語BERT
学習データ　JSICK・JSNLI
テストデータ　推論テストセット
（含意ラベル・推意ラベル）

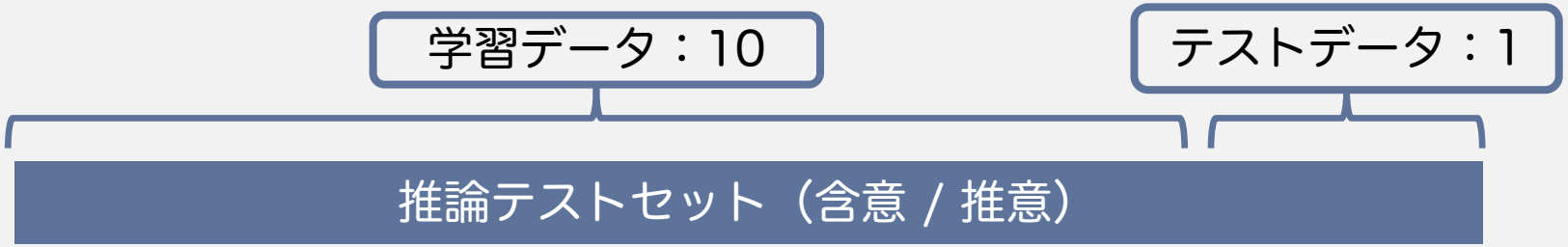
	含意ラベル		推意ラベル	
yes	653	33.78%	656	33.94%
no	660	34.14%	1092	56.49%
unknown	620	32.07%	185	9.57%
全体	1933		1933	

	JSICK		JSNLI	
	含意ラベル	推意ラベル	含意ラベル	推意ラベル
yes	62.48%	62.50%	70.29%	69.97%
no	9.39%	9.07%	38.03%	30.95%
unknown	28.39%	22.70%	15.48%	14.59%
全体	33.42%	28.50%	41.70%	42.63%

5. 評価

モデル　日本語BERT

学習方法　10分割交差検証



	含意_テスト		推意_テスト	
yes	57	32.39%	57	32.39%
no	55	31.25%	102	57.95%
unknown	64	36.36%	17	9.66%
全体	176		176	

	含意_学習		推意_学習	
	含意_テスト	推意_テスト	含意_テスト	推意_テスト
yes	70.18%	70.18%	73.68%	73.68%
no	72.73%	74.55%	83.64%	81.37%
unknown	81.25%	20.31%	20.31%	64.71%
全体	75.00%	53.41%	57.39%	77.27%

5. 評価

含意

文ペア	実験結果	正解
T: 会計検査院の調べでは、市では２０１１年１０月からの１年間、計１６４件の土木工事発注に対し８３件で入札が成立せず、不調率は５０．６％と被災自治体で最も高かった。 H: 会計検査院の調べでは、市では２０１１年１０月からの１年間、１６９件以上の土木工事発注があった。	yes	unknown
T: そこには、ユダヤ人のきよめのならわしに従って、それぞれ四、五斗もはいる石の水がめが、六つ置いてあった。 H: そこには置いてあった石の水がめはちょうど五つだった。	unknown	no

推意

文ペア	実験結果	正解
T: これはつまり6人もの子供が生まれたら少なくとも4人は次の世代として生き残るということです H: これはつまり7人以上の子供が生まれたら少なくとも4人は次の世代として生き残るということです	unknown	yes

6. おわりに

本発表

① 数量表現コーパス：
- 供述調書・NPCMJの数量表現を含む文を用いて高度な意味アノテーションを付与

② 推論テストセットの構築：
- 各文ペアに含意ラベルと推意ラベルを付与

③ ベースライン実験：
- 日本語BERTを用いてベースライン実験

今後について

① 数量表現コーパス・推論データセットの拡張

② 構築したデータセットを用いて含意関係認識システムの分析・評価を進める

③ 数量表現コーパスとそれに基づく推論データセットは研究利用可能な形式で公開予定

参考文献・URL

[1] Stephen Levinson. Pragmatics. Cambridge University Press, 1983. (S. レヴィンソン『英語語用論』, 安井 稔・ 奥田夏子訳, 研究社出版, 1990).

[2] NINJAL. NINJAL Parsed Corpus of Modern Japanese. (Version 1.0). Technical report, National Institute for Japanese Language and Linguistics, 2016. https://npcmj.ninjal.ac.jp/.

[3] Aakanksha Naik, Abhilasha Ravichander, Norman Sadeh, Carolyn Rose, and Graham Neubig. Stress test evaluation for natural language inference. In Proceedings of the 27th International Conference on Computational Linguistics, pp. 2340–2353, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics.

[4] Nelson F. Liu, Roy Schwartz, and Noah A. Smith. Inoculation by fine-tuning: A method for analyzing challenge datasets. 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), pp. 2171–2179, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[5] Paloma Jeretic, Alex Warstadt, Suvrat Bhooshan, and Adina Williams. 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] Ai Kawazoe, Ribeka Tanaka, Koji Mineshima, and Daisuke Bekki. An inference problem set for evaluating semantic theories and semantic processing systems for japanese. In New Frontiers in Artificial Intelligence, pp. 58–65, 2017.

[7] 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, 2015.

[8] 吉越卓見, 河原大輔, 黒橋禎夫. 機械翻訳を用いた自然言語推論テータセットの多言語化. 第244回自然言語処理研究会, 2020.

[9] Marco Marelli, Stefano Menini, Marco Baroni, Luisa Bentivogli, Raffaella Bernardi, and Roberto Zamparelli. A SICK cure for the evaluation of compositional distributional semantic models. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pp. 216–223, 2014.

[10] 谷中瞳, 峯島宏次. JSICK: 日本語構成的推論・類似度データセットの構築. 第35回人工知能学会全国大会, 2021.

[11] Yuta Hayashibe. Japanese realistic textual entailment corpus. In Proceedings of the 12th Language Resources and Evaluation Conference, pp. 6827–6834, 2020.

[12] 成澤克麻, 渡邉陽太郎, 水野淳太, 岡崎直観, 乾健郎. 数量表現を伴う文における含意関係認識の課題分析. 言語処理学会第18 回年次大会発表論文集, 2012.

[13] 飯田隆. 日本語と論理. NHK 出版, 2019.

[14] 奥津敬一郎. 拾遺日本文法論. ひつじ書房, 1996.

[15] 岩田一成. 日本語数量詞の諸相. くろしお出版, 2013.

[16] 法務総合研究所. 事件記録教材: 法科大学院教材. 第15号 (窃盗被疑事件). 法曹会, 2014.

[17] 法務総合研究所. 事件記録教材: 法科大学院教材. 第15号 (暴行機凝事件). 法曹会, 2014.

[18] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional trans- formers 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 Pa- pers), 2019.