# ファクトチェック支援のための含意関係認識システム

栗原健太郎
早稲田大学基幹理工学部
kkurihara@akane.waseda.jp

河原大輔
早稲田大学基幹理工学部
dkw@waseda.jp

## 1 はじめに

2016 年アメリカ大統領選挙以降、**フェイクニュース**という言葉が世界中に浸透している。フェイクニュースとは虚偽の情報で作られたニュース、すなわちデマ情報を意味する言葉であり、フェイクニュースという言葉の拡大は、政治家の発言や現地メディアの報道の真実性に関する議論を勃発させていた。しかし、それら発言や報道が、時として大統領選挙の結果などの重要な局面で大きな影響を与えうる。故に、報道機関では不正のない正当な情報の提供のために、発信する情報の真実性の向上と外部の言説の適切な利用を目的とした**ファクトチェック**の需要が高まっている。

しかしファクトチェックは、昨今のニュースの情報源の膨大さ故に大変な労力を要する作業である。日本においても、InFact[1)]や毎日新聞などのジャーナルが積極的なファクトチェックを行っているが、ファクトチェックの自動化が実現していない現状においては、その作業は人手や時間のかかる作業となっている。

本研究では、含意関係認識を活用したファクトチェック支援システムを考える。システムの概略図を図 1(a) に示す。本システムは以下のフローで構成される。

1. 疑義言説について関連する文書集合を取得
2. 疑義言説とその関連文書の含意関係認識

ここで疑義言説とは、ファクトチェックの対象となる真実かどうか疑わしい言説である。本研究では疑義言説を 1 文で表せるものとする。

本研究では、上記フローのうち、2 の含意関係認識のステップ (図 1(b)) の実現を目指す。このステップは疑義言説と関連文書との含意関係認識であるが、既存の含意関係認識リソースは文ペアを対象にしており、これらを利用することが難しい。そこで、本研究では、このステップを実現するための要素技術として、疑義言説 1 文と関連文書中の 1 文との含意関係認識を行うシステム (図 1(c)) を構築する。このシステムを NLIFC (NLI system for FactCheck) と呼ぶ。

NLIFC の訓練・評価のためにデータセットを構築した。これを FCSNLI (FactCheck Sentence NLI データセット) と呼ぶ。SNLI [1] や MultiNLI [2] などの既存の含意関係認識データセットでは、文ペアが contradiction (矛盾)", "entailment (含意)", "neutral (中立)" のいずれかのラベルをもち、2 文間には何らかの関係性がある。しかし、ファクトチェックにおいては、疑義言説と関連文書中の 1 文がまったく関係ないことも多い。そこで、FCSNLI のラベルは、上記 3 ラベルに "unrelated (無関係)" を加えた 4 ラベルとする。訓練データは、既存 3 ラベルの識別性能向上を目的として SNLI の和文翻訳データセットである JSNLI [3]、"unrelated" ラベルの識別性能向上を目的としてニュース記事のコーパス、数値表現の識別性能向上を目的とした擬似データの 3 種類から構成する。評価データは、ファクトチェック・イニシアティブ (FIJ) が提供する疑義言説データベースに掲載されている疑義言説を基に作成する。

NLIFC のモデルは文脈言語モデル BERT [4] を用いる。実験の結果、FCSNLI の 4 ラベルにおける F1 値は 0.653、"unrelated" 以外の 3 ラベルのみにおける F1 値は 0.538 となった。本研究における提案手法および構築したデータセットは、含意関係認識システムを活用したファクトチェック支援の第一歩になると考えている。

## 2 関連研究

田上ら [5] はファクトチェックすべきニュース記事 (要検証記事) の探索支援に関する研究を行っている。具体的には、ファクトチェックをするべきであるニュース記事の選別のための端緒情報の特定、および特定した端緒情報を利用したニュース記事の検証必要度のランク付けによる要検証記事の探索支援の仕組みを構築している。本稿では、このような仕組みによっ

1） https://infact.press/

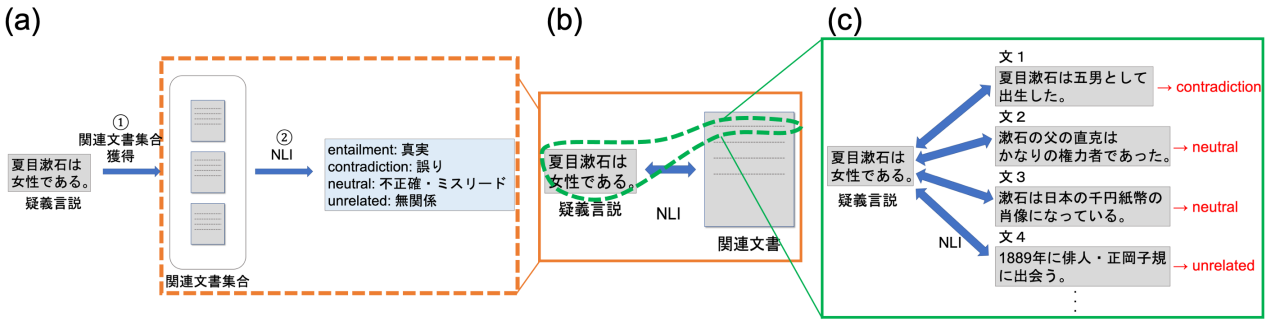


図 1　ファクトチェック支援システムの概略図

て抽出された要検証記事のファクトチェックを図 1 のフローで支援することを想定し、疑義言説と関連文書の含意関係認識システム構築の前段階として、疑義言説と関連文書中の文との含意関係認識システムを提案する。

## 3　ファクトチェック支援のための含意関係認識システム

ファクトチェック支援のための含意関係認識システムとして NLI system for FactCheck (NLIFC) を提案する。本モデルは図 1(c)、すなわち疑義言説と関連文書中の文の 2 文間の含意関係認識を行うシステムである。本研究では、疑義言説は 1 文で表されるものを扱うので**疑義言説文**、関連文書中の文を**リソース文**と呼ぶ。

提案モデルは BERT (Bidirectional Encoder Representations from Transformers) [4] をベースとする。本研究では BERT 日本語 Pretrained モデル [6] を使用し、ファクトチェック支援への活用に向けた含意関係認識タスクで fine-tuning する。本タスクにおける fine-tuning のために、4 節で述べる FCSNLI 訓練データを用いる。入力はリソース文と疑義言説文の 2 文であり、出力ラベルは “contradiction”, “entailment”, “neutral”, “unrelated” の 4 ラベルのいずれかである。

リソース文と疑義言説文は形態素解析器 Juman++[2) ] で分かち書きを行ったトークン列を入力とし、リソース文と疑義言説文の間に [SEP] トークンを追加している。

## 4　ファクトチェック支援に向けた含意関係認識データセットの構築

NLIFC の訓練および評価を行うデータセットとして、FCSNLI を構築する。本データセットのラベルは 1 節で述べたとおり、“contradiction”, “entailment, “neutral”, “unrelated” の 4 ラベルとした。

### 4.1　訓練データ

FCSNLI 訓練データは以下の 3 つのデータセットから構成する。

- JSNLI (contradiction: 182964, entailment: 182583, neutral: 182467, unrelated: 0)
- unrelated データセット (contradiction: 0, entailment: 0, neutral: 0, unrelated: 180000)
- figure データセット (contradiction: 248556, entailment: 248556, neutral: 30360, unrelated: 0)

unrelated データセットは、JSNLI データに含まれていない unrelated ラベルのデータを追加する目的で構築したデータセットである。データのペアのうち一方は livedoor ニュースコーパスの文を使用し、他方には JSNLI データセットの文を使用した。figure データセットは、NLIFC の数値表現の認識性能向上を目的として作成した擬似データセットである。リソース文・疑義言説文の数値を変化させ、対応する正解ラベルを与えることで擬似データの生成を行った。figure データセットの例を付録の表 6 に示す。

### 4.2　評価データ

FCSNLI 評価データはファクトチェック・イニシアティブ (FIJ) が提供する疑義言説データベース ClaimMonitor2 に掲載されている疑義言説のうち、Infact や毎日新聞などのジャーナルが検証を行った言説を基に構築した。データセットの構築は以下の 2 点に留意した上で人手で行った。

1. 人には容易に含意関係認識が可能となるようなデータセット
2. 既存の NLI データセットよりも平均文字数が多くなるような文のペアのデータセット

これらは、既存の NLI データセットにおける以下の特

2)　http://nlp.ist.i.kyoto-u.ac.jp/?JUMAN%2B%2B

| 正解ラベル | リソース文 | 疑義言説文 |
|---|---|---|
| contradiction | 東京オリンピックの会場となる国立競技場の整備関連にかかった費用は 1569 億円である。 | 東京オリンピックの会場となる新国立競技場の整備関連にかかった費用は 2500 億円に達している。 |
| unrelated | 東京は日本の首都であるがゆえに、日本の主要機能が集約されているが、これが人口集中を招く原因となっている。 | 東京オリンピックの会場となる新国立競技場の整備関連にかかった費用は 2500 億円に達している。 |
| neutral | 中国から送られてくる種子はジャイアントホグウィードであるという説があるが、正確な情報ではなく、別の種類の種子であるとも考えられている。 | 中国から送られてくる種子はジャイアントホグウィードである。 |
| unrelated | 中国は広大な草原や砂漠、山、湖、川を有する大国であり、人口も非常に多い。 | 中国から送られてくる種子はジャイアントホグウィードである。 |

**表 1**　FCSNLI 評価データの例

| | JSNLI (test) | FCSNLI (dev, test) |
|---|---|---|
| 前提文・リソース文 | 28.8 | 58.8 |
| 仮説文・疑義言説文 | 16.0 | 42.8 |

**表 2**　JSNLI・FCSNLI における文の平均文字数

徴との差別化を図ったものである。

1. ペアに無関係のものがほとんど含まれていない。
2. 文の平均文字数が短い。

FCSNLI 評価データの例を表 1 に示す。データ数は 開発 (dev) データが 305 ペア、テスト (test) データが 200 ペアとなっており、ラベルの比率は unrelated: 既存 3 ラベル=4:1 である。文の平均文字数は表 2 のように JSNLI と比較して多くなった。

# 5　含意関係認識システムの評価実験

## 5.1　実験設定

NLIFC の構築にあたって、BERT の学習の際のハイパーパラメータを付録の表 7 に示す。また、NLIFC の性能評価のため、figure データセットを除いて fine-tuning したモデル (以下 NLIFC - figure) と、JSNLI データセットのみを用いて fine-tuning したモデル (以下 NLIFC - unrelated - figure) も同様の設定で構築する。

モデルの性能評価には FCSNLI と JSNLI の 2 つの評価データを使用した。このうち JSNLI は unrelated 以外の 3 ラベルで構成されているデータセットであるため、NLIFC と NLIFC - unrelated - figure の比較を通じて unrelated の学習が与える既存 3 ラベルの識別性能への影響の調査が可能となる。精度の測定に関して、JSNLI データセットについては正答率をスコアとし、FCSNLI データセットについてはマイクロ平均により算出した F 値をスコアとしており、3 ラベルデータのスコア算出は unrelated データ以外について行った。

## 5.2　結果・考察

FCSNLI の全てのデータで訓練した結果を表 3 の NLIFC の行に示す。提案手法は JSNLI データセットにおいて、精度 0.922 を達成しており、JSNLI モデルのスコアとわずか 0.5%差であることから、unrelated の学習に伴う 3 ラベルの識別性能の低下は生じていないことを示している。また、FCSNLI - test の 4 ラベルデータで評価した結果は 3 ラベルデータで評価した結果と比較して 11.5 ポイント上回っている。これは NLIFC の unrelated ラベルの識別性能が他の 3 ラベルの識別性能よりも性能が高いことを示している。一方で 3 ラベルデータにおける 0.538 というスコアは、JSNLI タスクで高精度を達成している BERT も、ファクトチェック支援などの実応用における含意関係認識に関しては性能が不十分であることを示している。

NLIFC の予測例を表 4 に示す。(1) のようにリソース文の肯定表現と疑義言説の否定表現を正しく認識できた例が得られた一方で、(2)(3) のように “entailment” と判定してしまう誤り例も得られた。これはリソース文と疑義言説文とで一致する表現が多いことから否定表現の認識が適切に行われず “entailment” と判定してしまったと考えられる。次に (4) のようにリソース文あるいは疑義言説文が長文である場合の誤り例も多く見られた。これは長文における文脈の考慮が正確にできていないことに起因していると考えられる。(5) はリソース文の前半の「新型コロナ〜論文が存在する一方で」と疑義言説文との含意関係は “entailment”、リソース文の後半の節「WHO は〜続いている。」と疑義言説文との含意関係は “contradiction” となるという問題であり、このような節ごとに異なる含意関係を持つデータに関しても誤った判定をしていた。

| 手法 | JSNLI | FCSNLI-test (4 ラベル) | FCSNLI-test (3 ラベル) | FCSNLI-dev (4 ラベル) | FCSNLI-dev (3 ラベル) |
|---|---|---|---|---|---|
| NLIFC | 0.922 | 0.653 | 0.538 | 0.472 | 0.407 |
| NLIFC - figure | 0.923 | 0.581 | 0.486 | 0.495 | 0.438 |
| NLIFC - unrelated - figure | 0.927 | - | - | - | - |

**表 3**　評価結果

| | 正解ラベル | 予測 (NLIFC) | リソース文 | 疑義言説文 |
|---|---|---|---|---|
| (1) | contradiction | contradiction | 大村知事はコロナ感染者の個人情報を誤掲載したことを受けて、テレビ番組を通じて再発防止を述べるとともに、謝罪を行った。 | 大村知事はコロナ感染者の個人情報を誤って掲載したことに関して、一切謝罪をしていない。 |
| (2) | contradiction | entailment | 高知県で初めて新型コロナウイルスの感染が確認された３０代の女性は、その後も看護師を続けており、解雇されていない。 | 高知県で初めて新型コロナウイルスの感染が確認された３０代の看護師の女性は、その後看護師を解雇された。 |
| (3) | neutral | entailment | 世田谷一家殺人事件の犯人が日本人ではないという根拠のない言説がある。 | 世田谷一家殺人事件の犯人は日本人ではない。 |
| (4) | neutral | unrelated | 在日米国大使館の注意喚起によれば、現在の日本の医療制度に信頼を置いているものの、今後の数週間でどの程度機能するのか予測困難で、感染者増加によって米国市民が日本の医療を受けることができなくなる可能性を示唆しているが、「日本の医療は崩壊する」と断定はしていない。 | 在日米国大使館は「近いうちに日本は医療崩壊する」と警告した。 |
| (5) | neutral | unrelated | 新型コロナウイルスは高温多湿と紫外線に弱いとする論文が存在する一方で、WHO は「新型コロナウイルスはどれだけ日光があろうが、気温が高かろうが感染する」と警告を出しており、それを裏付ける論文が存在する上、高温多湿の東南アジアで実際に感染拡大が続いている。 | 新型コロナウイルスは、高温多湿と紫外線に弱い。 |

**表 4**　NLIFC の予測例

| | 正解ラベル | 予測 (NLIFC) | 予測 (NLIFC - figure) | リソース文 | 疑義言説文 |
|---|---|---|---|---|---|
| (Ⅰ) | contradiction | contradiction | neutral | 看護師によると、PCR 検査に用いる綿棒の長さはインフルエンザの検査に用いるものとほとんど同じ長さであり、医学雑誌のウェブサイトに掲載されている動画によると、検査の際の綿棒を回す回数は５回程度で、１０回とは大きく異なる。 | PCR 検査は、インフルエンザの検査に用いる綿棒の２倍の長さの綿棒を鼻に入れて１０回ほど回すというもので、口でも同様の作業を行う。 |
| (Ⅱ) | neutral | unrelated | neutral | PCR 検査で陽性と判明した女性看護師が勤務させられていた大阪生野区の病院における感染者数は数日間の累計 120 人であった。 | 新型コロナウイルスへの感染が確認された看護師の勤務が原因で、大阪生野区の病院では二日間に 120 人以上の大規模感染が起きた。 |

**表 5**　NLIFC と NLIFC - figure の予測例

### 5.3　Ablation 実験

NLIFC 構築において fine-tuning に用いた figure データセットが与えた影響に関して考える。NLIFC から figure データセットを除いて学習したモデル (NLIFC - figure) の評価結果を表 5 に示す。開発データ同士での比較ではほとんどスコアに差はなかったものの、テストデータにおいては NLIFC が NLIFC - figure のスコアを、4 ラベル評価・3 ラベル評価の両方で上回った。この結果は figure データセットの有効性を示している。各モデルの出力例を表 5 に示す。(Ⅰ) の他 3 件の正解ラベルが “contradiction” の数値表現を含むデータに関して同様に NLIFC のみが正解できた一方、(Ⅱ) の他 2 件の正解ラベルが “neutral” の数値表現を含むデータに関しては NLIFC - figure のみが正解した。これは、figure データセットの多くが “contradiction” と “entailment” の 2 つのラベルから構成されるため、neutral の数値表現に関しての学習が不十分であることが考えられる。

## 6　おわりに

本研究では、疑義言説と関連文書との含意関係認識システム構築の前段階として、疑義言説と関連文書中の文との含意関係認識システムとして NLIFC を提案した。また、ファクトチェックへの応用を想定した含意関係認識システムの評価指標として FCSNLI データセットを構築した。

実験結果は、FCSNLI タスクが JSNLI タスクと比較して難易度の高いタスクであることを示している。一方で、提案手法は FCSNLI タスクで F1 値 0.653 を達成しており、今後のファクトチェック支援の第一歩になると考えている。今後は、MultiNLI などのドメインのカバレッジが広い NLI データセットの和文翻訳データセットを用いた FCSNLI データセットの拡張による含意関係認識システムの性能向上を図るとともに、データセット拡張の自動化も検討していく。また、図 1(b) のような疑義言説と複数文から成る関連文書の含意関係認識システムの実現が今後の課題である。

# 参考文献


[1]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.

[2]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.

[3]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データセットの多言語化. 研究報告自然言語処理（NL）, pp. 1 – 8, 2020.

[4]Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 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)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[5]田上翼, 朝野広樹, 楊井人文, 山下亮, 小宮篤史, 藤村厚夫, 町野明徳, 乾健太郎. ファクトチェックを必要とするニュース記事の探索の支援. 言語処理学会 第 24 回年次大会, pp. 404 – 407, 2018.

[6]柴田知秀, 河原大輔, 黒橋禎夫. Bert による日本語構文解析の精度向上. 言語処理学会 第 25 回年次大会, pp. 205 – 208, 2019.

# A 付録

| 正解ラベル | リソース | 疑義言説 |
|---|---|---|
| contradiction | 山田君は 1 月 15 日から 3 月 14 日までの 2 ヶ月間、合宿で沖縄県に滞在していた。 | 山田君は 3 月 15 日に沖縄に滞在していた。 |
| entailment | 日本の人口 130000000 人の 50%が男性である。 | 日本の男性の人口は 14040000 人以上である。 |
| neutral | 瀬川君は 1 月 15 日から 6 月 14 日までの 5 ヶ月間、出張で京都府に滞在していた。 | 瀬川君は 7 月 1 日に神奈川県に滞在していない。 |
| contradiction | 日本の 140000000 人の 50%が女性である。 | 日本の女性の人口は 88900000 人より多い。 |

**表 6** figure データセットの例

| | |
|---|---|
| max-seq-length | 128 |
| train-batch-size | 32 |
| learning-rate | 2e-5 |
| epoch | 3 |

**表 7** ハイパーパラメータの設定値