

NTCIR 14

QA Lab Polilnfo
第4回説明会

1901/1902会議室@NII
2018年10月29日15:00-17:30



NTCIR-14 QA Lab PolilInfo

<https://poliinfo.github.io/>

Organizers:

木村泰知（小樽商科大学, 理化学研究所AIP）

渋木英潔（横浜国立大学）

高丸圭一（宇都宮共和大学）

内田ゆづ（北海学園大学）

乙武北斗（福岡大学）

石下円香（国立情報学研究所）

阪本浩太郎（横浜国立大学, 国立情報学研究所）

三田村照子（カーネギーメロン大学）

神門典子（国立情報学研究所, 総合研究大学院大学）

Advisers:

森辰則（横浜国立大学）

湯淺墾道（情報セキュリティ大学院大学）

関根聰（理化学研究所AIP）

乾健太郎（東北大学, 理化学研究所AIP）

第4回説明会の概要

- Formal Run の Training data

- ▶ Segmentation task
- ▶ Summarization task
- ▶ Classification task

- Formal Run に向けて

- ▶ 配布データについて
- ▶ みなさんから一言
- ▶ 今後の予定

- 懇親会

Task description

1. Segmentation Task

- ▶ 入力：質問と答弁の「発言の引用 の対(つい)」とその発言が含まれる「議会会議録」
- ▶ 出力：引用を理解するために読むべき発言の範囲（「開始行」「終了行」）

2. Summarization Task

- ▶ 入力：議会会議録に含まれる「発言」と要約の「制限字数」と「議会会議録」
- ▶ 出力：発話者の意図が伝わる「要約」

3. Classification Task

- ▶ 入力：政策が記述された「課題文」と議会会議録中の「発言文（一文のみ）」
- ▶ 出力：3値に分類「事実検証可能な根拠を伴う賛成」「事実検証可能な根拠を伴う反対」「その他」
 - ✓ 3つのサブクラス
 - 「関連（有り、無し）」
 - 「事実検証（可能、不可能）」
 - 「立場（支持、不支持、どちらでもない）」

Segmentation

Segmentation task 開始行について

●開始行の決定方法

▶ まず、開始行を説明する前に「都議会会議録」と「都議会だより」の関係について説明します

The screenshot shows the official website of the Tokyo Metropolitan Assembly (東京都議会). The header includes links for site map, mobile version, English, font size (拡大, 標準), search bar, and navigation menu items like Assembly Introduction, Member Introduction, Meeting Results and Minutes, Observation, and Survey. Below the header, a breadcrumb trail indicates the page path: トップ > 都議会だより > 297号 > 代表質問(吉田信夫). A large banner features the text "防災、放射能対策を強化せよ" and "暮らしと福祉を守る手立てを". Below the banner is a video player with a play button, a progress bar at 00:00 / 00:00, and the speaker's name "吉田信夫 (日本共産党)". A transcription table follows, with the first row labeled "防災". The table contains several entries with numbers in parentheses, such as "(1) 東京の津波対策の遅れをどう認識し、対応するのか。 (2) 耐震化されていない河川堤防は、東部地域で68キロ残る。強化策を。 (3) 超高層マンションの長周期地震動対応に支援を。 (4) また、室内の安全対策を図るべき。" and other responses from various officials.

| 防災 |
|---|
| (1) 東京の津波対策の遅れをどう認識し、対応するのか。 (2) 耐震化されていない河川堤防は、東部地域で68キロ残る。強化策を。 (3) 超高層マンションの長周期地震動対応に支援を。 (4) また、室内の安全対策を図るべき。 |
| 知事 (1) 新たな被害想定の見直しを進めており、津波対策に万全を期していく。 |
| 東京都技監 (2) 技術検証委員会等で耐震性向上策を検討。これを踏まえ新整備計画を策定。 |
| 都市整備局長 (3) 対策を国に要望した。 |
| 消防総監 (4) 専門委員会を設置し審議した。今後も、総合的な室内安全対策を推進する。 |

Segmentation task 開始行について

● 開始行の決定方法

- ▶ 「東京都議会だより」と「都議会会議録」の関係について

The screenshot shows a webpage from the Tokyo Metropolitan Assembly website. At the top, there's a blue header with the assembly's name in Japanese and English, and links for site map, mobile view, and English version. Below the header, a navigation bar includes links for assembly introduction, member introduction, meeting results and records, tours, and investigations. The main content area shows a breadcrumb trail: トップ > 都議会だより > 297号 > 代表質問(吉田信夫). The main content is a speech transcript:

防災、放射能対策を強化せよ
暮らしと福祉を守る手立てを

Below the transcript is a video player interface showing a progress bar and the time 00:00 / 00:00. To the right, a yellow box identifies the speaker as 吉田信夫 (日本共産党).

Annotations on the left side categorize parts of the speech:

- Main topic: Points to the first sentence "防災、放射能対策を強化せよ".
- Sub topic: Points to the word "防災".
- 質問要約 (Question Summary): Points to the question part of the transcript.
- 答弁者 (Speaker): Points to the speaker's name "吉田信夫 (日本共産党)".

Annotations on the right side categorize parts of the speech:

- 質問者 (Questioner): Points to the speaker's name "吉田信夫 (日本共産党)".
- 答弁要約 (Answer Summary): Points to the summary part of the transcript.

The transcript itself is divided into sections by the speakers:

- 防災 (Disaster Prevention): (1) 東京の津波対策の遅れをどう認識し、対応するのか。 (2) 耐震化されていない河川堤防は、東部地域で68キロ残る。強化策を。 (3) 超高層マンションの長周期地震動対応に支援を。 (4) また、室内の安全対策を図るべき。
- 知事 (Governor): (1) 新たな被害想定の見直しを進めており、津波対策に万全を期していく。
- 東京都技監 (Tokyo Metropolitan Technical Advisor): (2) 技術検証委員会等で耐震性向上策を検討。これを踏まえ新整備計画を策定。
- 都市整備局長 (City Construction Bureau Director): (3) 対策を国に要望した。
- 消防総監 (Fire Commissioner): (4) 専門委員会を設置し審議した。今後も、総合的な室内安全対策を推進する。

Segmentation のJsonフォーマット(Formal run バージョン)

● Segmentation Task

- ▶ 質問と答弁の要約を与えます
- ▶ 質問の範囲と答弁の範囲を答えてもらいます

- ✓ 質問の開始行
- ✓ 質問の終了行
- ✓ 答弁の開始行
- ✓ 答弁の終了行

| 表記 | 説明 |
|----------------------|-----------------------------|
| ID | 識別子 タスク名-開催年-言語-タスク種別-連番 |
| Prefecture | 都道府県名 |
| Date | 日付 |
| Meeting | 会議名 (議会だよりの記載) |
| MainTopic | メイントピック |
| SubTopic | サブトピック |
| QuestionSpeaker | 質問者 |
| QuestionSummary | 質問の要約 |
| AnswerSpeaker | 答弁者 |
| AnswerSummary | 答弁の要約 |
| QuestionStartingLine | 質問の開始行 |
| QuestionEndingLine | 質問の終了行 |
| AnswerStartingLine | 答弁の開始行 |
| AnswerEndingLine | 答弁の終了行 |

Segmentation のJsonフォーマット(Formal run バージョン)

```
{  
    "ID": "Segmentation-2018-JA-FormalSample-00001",  
    "Prefecture": "東京都",  
    "Date": "23-6-23",  
    "Meeting": "平成23年_第2回定例会",  
    "MainTopic": "東京の防災力向上は待ったなし<br>被災地復興には総合的取組を",  
    "SubTopic": "都政運営",  
    "QuestionSpeaker": "村上英子（自民党）",  
    "QuestionSummary": "(1) 国難のもとにおける東京の役割について所見を。 (2) 「10年後の東京」計画改定の視点は。 (3) 都民への責任果たし得る財政運営を。 (4) 被災地復興支援と首都東京の防災力向上に今後どう取り組むのか、所見を。",  
    "AnswerSpeaker": "知事本局長",  
    "AnswerSummary": "(2) 高度防災機能備えた都市へ取組強化等で安全・安心社会をつくり、節電意識徹底等で環境と経済が両立した都市実現等。",  
    "QuestionStartingLine": 763,  
    "QuestionEndingLine": 793,  
    "AnswerStartingLine": 1240,  
    "AnswerEndingLine": 1243  
},
```

埋める

| 表記 | 説明 |
|----------------------|-----------------------------|
| ID | 識別子 タスク名-開催年-言語-タスク種別-連番 |
| Prefecture | 都道府県名 |
| Date | 日付 |
| Meeting | 会議名 (議会だよりの記載) |
| MainTopic | メイントピック |
| SubTopic | サブトピック |
| QuestionSpeaker | 質問者 |
| QuestionSummary | 質問の要約 |
| AnswerSpeaker | 答弁者 |
| AnswerSummary | 答弁の要約 |
| QuestionStartingLine | 質問の開始行 |
| QuestionEndingLine | 質問の終了行 |
| AnswerStartingLine | 答弁の開始行 |
| AnswerEndingLine | 答弁の終了行 |

Training dataset for segmentation

| | |
|----|-----|
| 質問 | 298 |
| 答弁 | 298 |

298セット

質問と答弁を対(つい)としているので
298セットとなる

Summarization

SummarizationのJsonフォーマット(Formal run バージョン)

● Summarization Task

- ▶ 都議会会議録データを利用します
 - ✓ Pref13_tokyo.json
- ▶ トピックについて
 - ✓ 「メイントピック」「サブトピック」の追加
- ▶ 会議録データの対象範囲を明記します
 - ✓ 「開始行」「終了行」の追加

| 表記 | 説明 |
|--------------|-----------------------------|
| ID | 識別子 タスク名—開催年—言語—タスク種別—連番 |
| Prefecture | 都道府県名 |
| Date | 日付 |
| Meeting | 会議名 (議会だよりの記載) |
| Speaker | 発言者 |
| StartingLine | 開始行 |
| EndingLine | 終了行 |
| Main topic | メイントピック |
| Sub topic | サブトピック |
| Summary | 要約文 |
| Length | 文字数 |
| Source | 原文 |

SummarizationのJsonフォーマット(Formal run バージョン)

{
 "ID": "Summarization-2018-JA-Dry-00002",

 "Prefecture": "東京都",

 "Date": "23-6-23",

 "Meeting": "平成23年第2回定例会",

 "Speaker": "石原慎太郎",

 "StartingLine": 541,

 "EndingLine": 550,

 "Main topic": " 東京の総合防災力を更に高めよ
環境に配慮した都市づくりを",

 "Sub topic": "原子力発電所",

 "Summary": " 原子力利用に当たって安全の確保は当然。信頼性失ってきた姿勢反省し再考すべき。
同時にエネルギー確保は国家・社会の存立に直結、戦略を速やかに構え直す必要あり。 ",

 "Length": "100字以内",

埋める

 "Source": "次いで、原子力発電の安全性についてであります、今回の福島原発の事故は、ただいまの質問にあつた四年前、柏崎原発が被災した際、東京都が国に安全性確保を要求した云々の話とは全く次元の違う問題であります。¥¥n先日、海江田経済産業大臣は、原発の短期的な安全については確認されたと表明しております。¥¥nみずから民主党政権が原発を管轄し、検査もしているのですから、安全性については国家に聞いたらいかがですか。¥¥n改めて申し上げることもなく、原子力の利用に当たっても、安全の確保は当然であります。¥¥nなぜフランスでできている安全管理が日本にできないのか。¥¥nインシデントすら隠して信頼性を失ってきたこれまでの姿勢反省し、再考すべきであると思います。¥¥n同時に、エネルギー確保は、国家、社会の存立に直結するがゆえに、エネルギー戦略も速やかに構え直す必要があります。¥¥n豊富な電力が安定的に供給されるからこそ、我が国を支えるものづくりがここまで発展できたわけでありまして、新しいITの技術が急速に進歩してまいりました。¥¥nかつて、完全な管理技術を前提とすれば、東京湾に立派な原子力発電をつくってもいいと発言したことがございますが、当然のことであります。¥¥n日本の頭脳、心臓部の将来を、一段の発展を考えたとき、その裏づけとなる電力をほかに依存し続けることは不合理だと考えます。¥¥n"

},

| 表記 | 説明 |
|--------------|-----------------------------|
| ID | 識別子 タスク名-開催年-言語-タスク種別-連番 |
| Prefecture | 都道府県名 |
| Date | 日付 |
| Meeting | 会議名 (議会だよりの記載) |
| Speaker | 発言者 |
| StartingLine | 開始行 |
| EndingLine | 終了行 |
| Main topic | メイントピック |
| Sub topic | サブトピック |
| Summary | 要約文 |
| Length | 文字数 |
| Source | 原文 |

Training dataset for summarization

| | |
|----|-----|
| 質問 | 298 |
| 答弁 | 298 |

596セット

質問と答弁を分けているので
Segmentationの2倍になる

Classification

ラベル付けについて

●目的

- ▶ 政治家の発言から「事実検証可能な根拠を伴う意見」を見つけることです。

●作業

- ▶ 「課題文」と「政治家の発言(一文のみ)」を読み、次の3つのラベルを付けてもらいます。

1. 「関連」 ... 有り or 無し
2. 「事実検証可能」 ... 可能 or 不可能
3. 「立場」 ... Support , Against or Other

ラベル付けについて

▶ 課題文とは

✓ 政治課題と意見が含まれており、「政治家の発言(一文のみ)」と比較する文のことです。

- 例 「築地市場を豊洲に移転するべきである」

▶ 関連(Relevance)とは

✓ 「政治家の発言(一文のみ)」と「課題文」との間にかかわりがあることです。

- 関連のある例 「新銀行東京や築地市場の移転問題は非の立場です。」
- 関連のない例 「お正月五日の築地の初競りで史上最高となる一億五千五百四十万で落札された大間マグロはブランド化成功的代表例であります」

▶ 事実検証可能(Fact-checkability)とは

✓ 「政治家の発言(一文のみ)」に施設名や日付や金額などの具体的な表現が含まれ、Fact Check(事実の検証)ができることです。

- 事実検証できる例 「このような中、東京都はこの八月三十日に、豊洲の土壤汚染対策工事として、ゼネコン系の三つのJVと合計約五百四十二億円の契約を交わしています。」
- 事実検証できない例 「私は、築地の場外市場のようなあの温かみのある、あの何ともいえないにぎわいをつくるのが、都の責任の一つであると考えます」

▶ 立場(Stance)とは

✓ 「課題文」に対する発言者の立ち位置です。Support(支持), Against(不支持), Neutral(中立,どちらでもない)から一つ選びます。

- Support は、課題文の内容を支持していること
 - Support の例 「豊洲は、新市場移転により千客万来施設ができるなど、今後、観光客の集客が大いに期待できるエリアであります。」
- Against は、課題文の内容を支持していないこと
 - Against の例 「豊洲の新市場予定地では108箇所で液状化が発生し、築地市場の移転先としてふさわしくないことが重ねて証明されました。」
- Otherl は、支持、不支持がはっきりしない場合や中立の場合
 - (発言者の言葉ではない) 引用や報告は意見ではありません。

関連、事実検証可能、立場のラベルをつける例 (1/4)

課題文「築地市場を豊洲に移転するべきである」と下記の発言文を読み、3つのラベルをつける

| 文ID | 発言文（文脈のない一文） ※ 連続した発言ではありません。 | 関連 | 事実検証可能 | 立場 |
|-----|--|----|--------|----|
| 1 | 豊洲は、新市場移転により千客万来施設ができるなど、今後、観光客の集客が大いに期待できるエリアであります。 | | | |
| 2 | 豊洲の新市場予定地では108箇所で液状化が発生し、築地市場の移転先としてふさわしくないことが重ねて証明されました。 | | | |
| 3 | 新銀行東京や築地市場の移転問題は非の立場です。 | | | |
| 4 | このような中、東京都はこの八月三十日に、豊洲の土壤汚染対策工事として、ゼネコン系の三つのJVと合計約五百四十二億円の契約を交わしています。 | | | |
| 5 | 豊洲新市場への移転を希望する事業者の不安を払拭すべく、移転資金や運転資金、移転後の新たな事業展開に必要な資金の手当てなど、経営支援策を講じてまいります。 | | | |
| 6 | お正月五日の築地の初競りで史上最高となる一億五千五百四十万で落札された大間マグロはブランド化成功の代表例であります | | | |
| 7 | 京築地域ではどのように三次医療が確保されているのかお聞かせ願います | | | |

関連、事実検証可能、立場のラベルをつける例 (2/4)

課題文「築地市場を豊洲に移転するべきである」と下記の発言文を読み、3つのラベルをつける

| 文ID | 発言文（文脈のない一文）※ 連続した発言ではありません。 | 関連 | 事実検証可能 | 立場 |
|-----|--|----|--------|----|
| 1 | 豊洲は、新市場移転により千客万来施設ができるなど、今後、観光客の集客が大いに期待できるエリアあります。 | 有 | | |
| 2 | 豊洲の新市場予定地では108箇所で液状化が発生し、築地市場の移転先としてふさわしくないことが重ねて証明されました。 | 有 | | |
| 3 | 新銀行東京や築地市場の移転問題は非の立場です。 | 有 | | |
| 4 | このような中、東京都はこの八月三十日に、豊洲の土壤汚染対策工事として、ゼネコン系の三つのJVと合計約五百四十二億円の契約を交わしています。 | 有 | | |
| 5 | 豊洲新市場への移転を希望する事業者の不安を払拭すべく、移転資金や運転資金、移転後の新たな事業展開に必要な資金の手当てなど、経営支援策を講じてまいります。 | 有 | | |
| 6 | お正月五日の築地の初競りで史上最高となる一億五千五百四十万で落札された大間マグロはブランド化成功の代表例であります | | 無 | |
| 7 | 京築地域ではどのように三次医療が確保されているのかお聞かせ願います | | 無 | |

関連、事実検証可能、立場のラベルをつける例 (3/4)

課題文「築地市場を豊洲に移転するべきである」と下記の発言文を読み、3つのラベルをつける

| 文ID | 発言文（文脈のない一文） ※ 連続した発言ではありません。 | 関連 | 事実検証可能 | 立場 |
|-----|--|----|--------|----|
| 1 | 豊洲は、新市場移転により千客万来施設ができるなど、今後、観光客の集客が大いに期待できるエリアあります。 | 有 | 可能 | |
| 2 | 豊洲の新市場予定地では108箇所で液状化が発生し、築地市場の移転先としてふさわしくないことが重ねて証明されました。 | 有 | 可能 | |
| 3 | 新銀行東京や築地市場の移転問題は非の立場です。 | 有 | 不可能 | |
| 4 | このような中、東京都はこの八月三十日に、豊洲の土壤汚染対策工事として、ゼネコン系の三つのJVと合計約五百四十二億円の契約を交わしています。 | 有 | 可能 | |
| 5 | 豊洲新市場への移転を希望する事業者の不安を払拭すべく、移転資金や運転資金、移転後の新たな事業展開に必要な資金の手当てなど、経営支援策を講じてまいります。 | 有 | 不可能 | |
| 6 | お正月五日の築地の初競りで史上最高となる一億五千五百四十万で落札された大間マグロはブランド化成功の代表例であります | 無 | 可能 | |
| 7 | 京築地域ではどのように三次医療が確保されているのかお聞かせ願います | 無 | 不可能 | |

関連、事実検証可能、立場のラベルをつける例 (4/4)

課題文「築地市場を豊洲に移転するべきである」と下記の発言文を読み、3つのラベルをつける

| 文ID | 発言文（文脈のない一文）※ 連続した発言ではありません。 | 関連 | 事実検証可能 | 立場 |
|-----|--|----|--------|---------|
| 1 | 豊洲は、新市場移転により千客万来施設ができるなど、今後、観光客の集客が大いに期待できるエリアであります。 | 有 | 可能 | 支持 |
| 2 | 豊洲の新市場予定地では108箇所で液状化が発生し、築地市場の移転先としてふさわしくないことが重ねて証明されました。 | 有 | 可能 | 不支持 |
| 3 | 新銀行東京や築地市場の移転問題は非の立場です。 | 有 | 不可能 | 不支持 |
| 4 | このような中、東京都はこの八月三十日に、豊洲の土壤汚染対策工事として、ゼネコン系の三つのJVと合計約五百四十二億円の契約を交わしています。 | 有 | 可能 | どちらでもない |
| 5 | 豊洲新市場への移転を希望する事業者の不安を払拭すべく、移転資金や運転資金、移転後の新たな事業展開に必要な資金の手当てなど、経営支援策を講じてまいります。 | 有 | 不可能 | どちらでもない |
| 6 | お正月五日の築地の初競りで史上最高となる一億五千五百四十万で落札された大間マグロはブランド化成功の代表例であります | 無 | 可能 | どちらでもない |
| 7 | 京築地域ではどのように三次医療が確保されているのかお聞かせ願います | 無 | 不可能 | どちらでもない |

関連、事実検証可能、立場 と Class の関係

| 関連 | 事実検証 | 立場 | Class |
|----|------|---------|------------------------|
| 有 | 可能 | 支持 | Fact-checkable Support |
| 有 | 可能 | 不支持 | Fact-checkable Against |
| 有 | 可能 | どちらでもない | Other |
| 有 | 不可能 | 支持 | Other |
| 有 | 不可能 | 不支持 | Other |
| 有 | 不可能 | どちらでもない | Other |
| 無 | 可能 | 支持 | Other |
| 無 | 可能 | 不支持 | Other |
| 無 | 可能 | どちらでもない | Other |
| 無 | 不可能 | 支持 | Other |
| 無 | 不可能 | 不支持 | Other |
| 無 | 不可能 | どちらでもない | Other |

課題文と文数と注釈者

14の課題文

Training data の異なり文数 約1万文
Training data の総文数 約3万文

注釈者総数

46人

異なり注釈者数

20人

| 番号 | 番号 | 課題文 | 注釈者 |
|----|----|---------------------------|-----|
| 1 | 02 | カジノを含む統合型リゾートを推進するべきである | 3 |
| 2 | 04 | 集団的自衛を認めるべきである | 3 |
| 3 | 05 | 八ヶ場ダムを建設を進めるべきである | 3 |
| 4 | 07 | 高齢者への医療助成を増やすべきである | 3 |
| 5 | 08 | 私学助成を推進するべきである | 3 |
| 6 | 09 | 中京都構想を推進するべきである | 5 |
| 7 | 10 | オスプレイを配備する | 3 |
| 8 | 12 | 特定秘密保護法案を進めるべきである | 3 |
| 9 | 13 | 道州制を導入するべきである | 3 |
| 10 | 16 | 子ども医療費を無料化にするべきである | 5 |
| 11 | 19 | 正規の教員を増やすべきである | 3 |
| 12 | 22 | 生活保護の基準額を引き下げるべきである | 3 |
| 13 | 25 | 東京にオリンピックを招致するべきである | 3 |
| 14 | 26 | 行政の判断で空き家を処理できるようにするべきである | 3 |

| 課題文 | 注釈者 | 関連有り | 関連無し | 事実検証可能 | 事実検証不可能 | 支持 | 不支持 | どちらでもない | FactCheckable-Support | FactCheckable-Against | Other | 文数 |
|---|-----|------|------|--------|---------|----|-----|---------|-----------------------|-----------------------|-------|------|
| 04-集団的自衛を認めるべきである-1515-A.json | A | | | | | | | | | | | 1515 |
| 04-集団的自衛権の行使を容認する-1515-B.json | B | | | | | | | | | | | 1515 |
| 04-集団的自衛権の行使を容認する-1515-C.json | C | | | | | | | | | | | 1515 |
| 16-子ども医療費を無料化にするべきである-241-D.json | D | | | | | | | | | | | 241 |
| 16-子ども医療費を無料化にするべきである-241-E.json | E | | | | | | | | | | | 241 |
| 16-子ども医療費を無料化にするべきである-241-F.json | F | | | | | | | | | | | 241 |
| 16-子ども医療費を無料化にするべきである-241-G.json | G | | | | | | | | | | | 241 |
| 16-子ども医療費を無料化にするべきである-241-H.json | H | | | | | | | | | | | 241 |
| 1 09-中京都構想の推進について-226-B.json | B | | | | | | | | | | | 226 |
| 1 09-中京都構想の推進について-226-I.json | I | | | | | | | | | | | 226 |
| 1 09-中京都構想の推進について-226-J.json | J | | | | | | | | | | | 226 |
| 1 09-中京都構想を推進するべきである-226-G.json | G | | | | | | | | | | | 226 |
| 1 09-中京都構想の推進について-226-C.json | C | | | | | | | | | | | 226 |
| 1 02-カジノを含む統合型リゾートの推進-1165-J.json | J | | | | | | | | | | | 1165 |
| 1 02-カジノを含む統合型リゾートの推進-1165-K.json | K | | | | | | | | | | | 1165 |
| 1 02-カジノを含む統合型リゾートの推進-1165-F.json | F | | | | | | | | | | | 1165 |
| 1 08-私学助成を推進するべきである-502-L.json | L | | | | | | | | | | | 502 |
| 1 08-私学助成を推進するべきである-502-G.json | G | | | | | | | | | | | 502 |
| 1 08-私学助成を推進するべきである-502-M.json | M | | | | | | | | | | | 502 |
| 1 10-オスプレイを配備するべきである-1174-J.json | J | | | | | | | | | | | 1174 |
| 1 10-オスプレイを配備するべきである-1174-L.json | L | | | | | | | | | | | 1174 |
| 1 10-オスプレイを配備するべきである-1174-A.json | A | | | | | | | | | | | 1174 |
| 1 22-生活保護の基準額を引き下げるべきである-827-N.json | N | | | | | | | | | | | 827 |
| 1 22-生活保護の基準額を引き下げるべきである-827-I.json | I | | | | | | | | | | | 827 |
| 1 22-生活保護の基準額を引き下げるべきである-827-A.json | A | | | | | | | | | | | 827 |
| 1 05-ハッ場ダムの建設について-886-I.json | I | | | | | | | | | | | 886 |
| 1 05-ハッ場ダムを建設を進めるべきである-886-F.json | F | | | | | | | | | | | 886 |
| 1 05-ハッ場ダムの建設について-886-H.json | H | | | | | | | | | | | 886 |
| 1 26-行政の判断で空き家を処理できるようにするべきである-298-N.json | N | | | | | | | | | | | 298 |
| 1 26-行政の判断で空き家を処理できるようにするべきである-298-O.json | O | | | | | | | | | | | 298 |
| 1 26-行政の判断で空き家を処理できるようにするべきである-298-P.json | P | | | | | | | | | | | 298 |
| 1 25-東京にオリンピックを招致するべきである-834-H.json | H | | | | | | | | | | | 834 |
| 1 25-東京にオリンピックを招致するべきである-834-M.json | M | | | | | | | | | | | 834 |
| 1 25-東京にオリンピックを招致するべきである-834-I.json | I | | | | | | | | | | | 834 |
| 1 19-正規の教員を増やすべきである-345-M.json | M | | | | | | | | | | | 345 |
| 1 19-正規の教員を増やすべきである-345-C.json | C | | | | | | | | | | | 345 |
| 1 19-正規の教員を増やすべきである-345-G.json | G | | | | | | | | | | | 345 |
| 1 12-特定秘密保護法案を進めるべきである-462-Q.json | Q | | | | | | | | | | | 462 |
| 1 12-特定秘密保護法案を進めるべきである-462-R.json | R | | | | | | | | | | | 462 |
| 1 12-特定秘密保護法案を進めるべきである-462-S.json | S | | | | | | | | | | | 462 |
| 1 13-道州制を導入するべきである-1152-G.json | G | | | | | | | | | | | 1152 |
| 1 13-道州制を導入するべきである-1152-B.json | B | | | | | | | | | | | 1152 |
| 1 13-道州制を導入するべきである-1152-H.json | H | | | | | | | | | | | 1152 |
| 1 07-高齢者への医療助成を増やすべきである-664-F.json | F | | | | | | | | | | | 664 |
| 1 07-高齢者の医療助成を増やすべきである-664-T.json | T | | | | | | | | | | | 664 |
| 1 07-高齢者への医療助成を増やすべきである-664-C.json | C | | | | | | | | | | | 664 |

Classification Training dataについて

14の課題文

異なり文数 約1万文
3~5人の注釈文の総数 約3万文

注意：注釈者への説明の際に、3つのサブクラスの用語を異なる表現（例えば、「根拠あり」と「事実検証可能」）で説明していますが、注釈付けするラベルの意図は、同じです。

| 課題文 | 注釈者 | 関連有り | 関連無し | 事実検証可能 | 事実検証不可能 | 支持 | 不支持 | どちらでもない | FactCheckable-Support | FactCheckable-Against | Other | 文数 |
|--|-----|------|------|--------|---------|-----|-----|---------|-----------------------|-----------------------|-------|------|
| 04-集団的自衛を認めるべきである-1515-A.json | A | 1426 | 89 | 151 | 1364 | 28 | 70 | 1417 | 0 | 1 | 1514 | 1515 |
| 04-集団的自衛権の行使を容認する-1515-B.json | B | 1502 | 13 | 1088 | 427 | 30 | 82 | 1403 | 29 | 75 | 1411 | 1515 |
| 04-集団的自衛権の行使を容認する-1515-C.json | C | 1509 | 6 | 234 | 1281 | 186 | 319 | 1010 | 17 | 36 | 1462 | 1515 |
| 16-子ども医療費を無料化にするべきである-241-D.json | D | 237 | 4 | 30 | 211 | 79 | 5 | 157 | 6 | 5 | 230 | 241 |
| 16-子ども医療費を無料化にするべきである-241-E.json | E | 235 | 6 | 70 | 171 | 21 | 5 | 215 | 2 | 0 | 239 | 241 |
| 16-子ども医療費を無料化にするべきである-241-F.json | F | 235 | 6 | 95 | 146 | 17 | 0 | 224 | 8 | 0 | 233 | 241 |
| 16-子ども医療費を無料化にするべきである-241-G.json | G | 236 | 5 | 39 | 202 | 68 | 0 | 173 | 8 | 0 | 233 | 241 |
| 16-子ども医療費を無料化にするべきである-241-H.json | H | 228 | 13 | 32 | 209 | 45 | 2 | 194 | 4 | 0 | 237 | 241 |
| 1) 09-中京都構想の推進について-226-B.json | B | 226 | 0 | 90 | 136 | 27 | 4 | 195 | 13 | 3 | 210 | 226 |
| 1) 09-中京都構想の推進について-226-I.json | I | 226 | 0 | 78 | 148 | 50 | 8 | 168 | 10 | 5 | 211 | 226 |
| 1) 09-中京都構想の推進について-226-J.json | J | 220 | 6 | 41 | 185 | 52 | 16 | 158 | 6 | 4 | 216 | 226 |
| 1) 09-中京都構想を推進するべきである-226-G.json | G | 226 | 0 | 7 | 219 | 38 | 5 | 183 | 2 | 0 | 224 | 226 |
| 1) 09-中京都構想の推進について-226-C.json | C | 224 | 2 | 10 | 216 | 31 | 5 | 190 | 5 | 0 | 221 | 226 |
| 1) 02-カジノを含む統合型リゾートの推進-1165-J.json | J | 1055 | 110 | 554 | 611 | 339 | 193 | 633 | 198 | 78 | 889 | 1165 |
| 1) 02-カジノを含む統合型リゾートの推進-1165-K.json | K | 1146 | 19 | 645 | 520 | 52 | 45 | 1068 | 23 | 15 | 1127 | 1165 |
| 1) 02-カジノを含む統合型リゾートの推進-1165-F.json | F | 1165 | 0 | 420 | 745 | 29 | 39 | 1097 | 2 | 4 | 1159 | 1165 |
| 1) 08-私学助成を推進するべきである-502-L.json | L | 499 | 3 | 221 | 281 | 160 | 15 | 327 | 80 | 10 | 412 | 502 |
| 1) 08-私学助成を推進するべきである-502-G.json | G | 502 | 0 | 188 | 314 | 109 | 3 | 390 | 22 | 2 | 478 | 502 |
| 1) 08-私学助成を推進するべきである-502-M.json | M | 406 | 96 | 219 | 283 | 100 | 3 | 399 | 46 | 2 | 454 | 502 |
| 1) 10-オスプレイを配備するべきである-1174-J.json | J | 933 | 241 | 691 | 483 | 34 | 314 | 826 | 11 | 124 | 1039 | 1174 |
| 1) 10-オスプレイを配備するべきである-1174-L.json | L | 1164 | 10 | 138 | 1036 | 8 | 145 | 1021 | 6 | 14 | 1154 | 1174 |
| 1) 10-オスプレイを配備するべきである-1174-A.json | A | 1172 | 2 | 84 | 1090 | 2 | 100 | 1072 | 0 | 11 | 1163 | 1174 |
| 1) 22-生活保護の基準額を引き下げるべきである-827-N.json | N | 824 | 3 | 216 | 611 | 5 | 70 | 752 | 0 | 14 | 813 | 827 |
| 1) 22-生活保護の基準額を引き下げるべきである-827-I.json | I | 562 | 265 | 724 | 103 | 8 | 101 | 718 | 5 | 76 | 746 | 827 |
| 1) 22-生活保護の基準額を引き下げるべきである-827-A.json | A | 201 | 626 | 131 | 696 | 0 | 6 | 821 | 0 | 0 | 827 | 827 |
| 1) 05-八ヶ場ダムの建設について-886-I.json | I | 882 | 4 | 508 | 378 | 244 | 148 | 494 | 102 | 99 | 685 | 886 |
| 1) 05-八ヶ場ダムを建設を進めるべきである-886-F.json | F | 799 | 87 | 261 | 625 | 117 | 48 | 721 | 15 | 7 | 864 | 886 |
| 1) 05-八ヶ場ダムの建設について-886-H.json | H | 882 | 4 | 164 | 722 | 169 | 85 | 632 | 39 | 16 | 831 | 886 |
| 1) 26-行政の判断で空き家を処理できるようにするべきである-298-N.json | N | 298 | 0 | 77 | 221 | 109 | 1 | 188 | 23 | 0 | 275 | 298 |
| 1) 26-行政の判断で空き家を処理できるようにするべきである-298-O.json | O | 245 | 53 | 209 | 89 | 85 | 26 | 187 | 57 | 15 | 226 | 298 |
| 1) 26-行政の判断で空き家を処理できるようにするべきである-298-P.json | P | 79 | 219 | 116 | 182 | 6 | 0 | 292 | 1 | 0 | 297 | 298 |
| 1) 25-東京にオンラインピックを招致するべきである-834-H.json | H | 820 | 14 | 93 | 741 | 228 | 19 | 587 | 35 | 2 | 797 | 834 |
| 1) 25-東京にオンラインピックを招致するべきである-834-M.json | M | 350 | 484 | 225 | 609 | 83 | 27 | 724 | 15 | 5 | 814 | 834 |
| 1) 25-東京にオンラインピックを招致するべきである-834-I.json | I | 646 | 188 | 387 | 447 | 142 | 27 | 665 | 57 | 13 | 764 | 834 |
| 1) 19-正規の教員を増やすべきである-345-M.json | M | 219 | 126 | 108 | 237 | 85 | 4 | 256 | 26 | 2 | 317 | 345 |
| 1) 19-正規の教員を増やすべきである-345-C.json | C | 280 | 65 | 174 | 171 | 97 | 5 | 243 | 27 | 1 | 317 | 345 |
| 1) 19-正規の教員を増やすべきである-345-G.json | G | 332 | 13 | 85 | 260 | 84 | 2 | 259 | 9 | 1 | 335 | 345 |
| 1) 12-特定秘密保護法案を進めるべきである-462-Q.json | Q | 284 | 178 | 179 | 283 | 34 | 179 | 249 | 15 | 45 | 402 | 462 |
| 1) 12-特定秘密保護法案を進めるべきである-462-R.json | R | 55 | 407 | 353 | 109 | 15 | 138 | 309 | 4 | 35 | 423 | 462 |
| 1) 12-特定秘密保護法案を進めるべきである-462-S.json | S | 430 | 32 | 54 | 408 | 11 | 31 | 420 | 0 | 0 | 462 | 462 |
| 1) 13-道州制を導入するべきである-1152-G.json | G | 1151 | 1 | 232 | 920 | 211 | 60 | 881 | 16 | 8 | 1128 | 1152 |
| 1) 13-道州制を導入するべきである-1152-B.json | B | 1146 | 6 | 574 | 578 | 81 | 28 | 1043 | 35 | 6 | 1111 | 1152 |
| 1) 13-道州制を導入するべきである-1152-H.json | H | 1145 | 7 | 128 | 1024 | 352 | 98 | 702 | 67 | 11 | 1074 | 1152 |
| 1) 07-高齢者への医療助成を増やすべきである-664-F.json | F | 560 | 104 | 131 | 533 | 47 | 0 | 617 | 6 | 0 | 658 | 664 |
| 1) 07-高齢者への医療助成を増やすべきである-664-T.json | T | 122 | 542 | 409 | 255 | 24 | 9 | 631 | 20 | 4 | 640 | 664 |
| 1) 07-高齢者への医療助成を増やすべきである-664-C.json | C | 547 | 117 | 291 | 373 | 300 | 13 | 351 | 93 | 7 | 564 | 664 |

Formal run で提供する Training dataset

1. Segmentation
 - ▶ Segmentation-2018-JA-FromalTraining.json
 - ▶ Pref13_tokyo.json
2. Summarization
 - ▶ Summarization-2018-JA-FormalTraining.json
 - ▶ Pref13_tokyo.json
3. Classification (14課題文、46ファイル)
 - ▶ Classification-2018-JA-FormalRun-Trainining-(整理用番号)-(課題文)-(文数)-(注釈者ID).json

Important dates

Feb 20, 2018: QALab-PolInfo Kickoff meeting in NII (room 1901, 1902)

Mar 20, 2018: NTCIR-14 Kickoff event in NII

Apr 19, 2018: 1st round table meeting in NII (room 1901, 1902)

May 31, 2018: 2nd round table meeting in NII (room 1208)

Jun 15, 2018: Dataset Release

Jul 30, 2018: Task Registration Due for Dry Run

Aug 6 - 9, 2018: Dry Run (Segmentation & Classification Tasks)

Aug 13 - 16, 2018: Dry Run (Summarization Task)

Aug 30, 2018: 3rd round table meeting in NII (room 2004 2005)

Oct 29, 2018: 4th round table meeting in NII

Nov 19, 2018: Task Registration Due for Formal Run (This is not required for Dry Run participants)

Nov 26 - 29, 2018: Formal Run (Segmentation & Classification Tasks)

Dec 3 - 6, 2018: Formal Run (Summarization Task)

Feb 1, 2019: Evaluation Result Release

Feb 1, 2019: Task overview paper release (draft)

Mar 15, 2019: Submission due of participant papers

May 1, 2019 Camera-ready participant paper due

Jun 2019: NTCIR-14 Conference & EVIA 2019 in NII, Tokyo

第4回説明会の概要

● Formal Run の Training data

- ▶ Segmentation task
- ▶ Summarization task
- ▶ Classification task

● Formal Run に向けて

- ▶ 配布データについて
 - ✓ Training data, Test data 以外に, Developing data に分けるべきですか？
- ▶ みなさんから一言
- ▶ 今後の予定

● 懇親会

Formal run の予定

| 月 | 火 | 水 | 木 | 金 | 土 | 日 |
|----|----|----|----|----|----|----|
| 26 | 27 | 28 | 29 | 30 | 1 | 2 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| 31 | 1 | 2 | 3 | 4 | 5 | 6 |

| Start | End | Segmentation | Summarization | Classification |
|-----------|-----------|--------------|---------------|----------------|
| 11/26 (月) | 11/29 (木) | E2E | | E2E |
| 12/03 (月) | 12/06 (木) | TO集計 | E2E | TO集計 |
| 12/10 (月) | 12/13 (木) | Check | Evaluation | Check |
| 12/17 (月) | 12/20 (木) | TO再集計 | TO集計 | TO再集計 |
| 12/24 (月) | 12/27 (木) | 公開予備 | 公開予備 | 公開予備 |

まとめ

- Formal Run の Training data

- ▶ Segmentation task
- ▶ Summarization task
- ▶ Classification task

- Formal Run に向けて

- ▶ 配布データについて

✓ Training data, Test data 以外に, Developing data に分けるべきですか？

● 説明会に参加者の意見を踏まえて、まず、説明会で提案したTraining dataを参加者にbitbucket経由で公開します。

- ▶ みなさんから一言

✓ (次のスライドに載せておきます)

- ▶ 今後の予定

- 懇親会

2018年10月29日の質疑まとめ

● 全体

- ▶ 質問：Dry runでは一つのタスクにしか参加していませんでしたが、Formal run から他のタスクに参加できますか？
- ▶ 回答：参加できます。ぜひ参加してください。
- ▶ 質問：1月中旬までにSystem Descriptionを欲しいと言いましたが、書く際に雛形をもらえないでしょうか？
- ▶ 回答：雛形をつくります。

● Segmentation について

- ▶ 質問：機械学習の手法を利用するには、データが少ないのでしょうか？
- ▶ 回答：ご指摘の通り、データが少ないのでですが、Formal runはこのデータで進めることになります。
- ▶ 質問：Json file の発言者が役職名になっているのはどうしてですか？
- ▶ 回答：議会だよりと会議録で表記が違うためです。「都議会だより」には、首長に代わって答弁する職員の名前は記載されておらず、役職名のみになっています。
- ▶ 質問：注釈は、2人でつけたようですが、開始行と終了行を合わせたのですか？
- ▶ 回答：その通りです。

● Summarization について

- ▶ 質問：ROUGEのスクリプトを公開してもらえないでしょうか？
- ▶ 回答：できるだけ早く参加者に公開します。

2018年10月29日の質疑まとめ

● Classificationについて

- ▶ 質問：(自分で会議録を収集して) 文脈を利用しても良いでしょうか？
- ▶ 回答：利用しても問題ありませんが、注釈者には（文脈を与える）一文のみの情報で判定してもらっています。
- ▶ 質問：「関連」の定義が難しいので、もう少し、詳しく記述してもらえないでしょうか？
- ▶ 回答：「関連」の定義は、注釈者に、定義文だけで説明するが難しいため、FAQ形式で例文を用いて説明しています。注釈者に渡している説明資料も公開します。
- ▶ 質問：Fact-checkableの意味をもう一度説明してください。
- ▶ 回答：Fact checkする価値があるものです（今すぐチェックするわけではありません）
- ▶ 質問：「事実検証可能」を判断するために、注釈者はGoogleでチェックしているのでしょうか？
- ▶ 回答：注釈者は、事実検証できる具体的な記述なのかを判断しているだけなので、Googleでチェックしていません。
- ▶ 質問：Trainingが3万というのは、1万文に3人が注釈付けしたという意味ですか？
- ▶ 回答：その通りです。
- ▶ 質問：14の課題文明はどのように決めたのですか？
- ▶ 回答：「○○について」などを含む表現が多い課題をピックアップしてから、意見が割れなさそうな課題を選択しました。
- ▶ 質問：判断対象となる文はどうように決めたのですか？
- ▶ 回答：grepでキーワードとなる単語を含む発言文を抽出しました。その後、30文字以上～200字未満の文のみを利用することとしました。
- ▶ 質問：「関連なし」は意見もなしになるのではないですか？
- ▶ 回答：その通りだと思いますが、注釈者が意見をついている場合もあるかもしれません。
- ▶ 質問：（評価が難しいと言ってましたが）評価方法をサブタスク化してはどうですか？
- ▶ 回答：次回反映することを考えます。