

NTCIR-15

QA Lab-PoliInfo-2

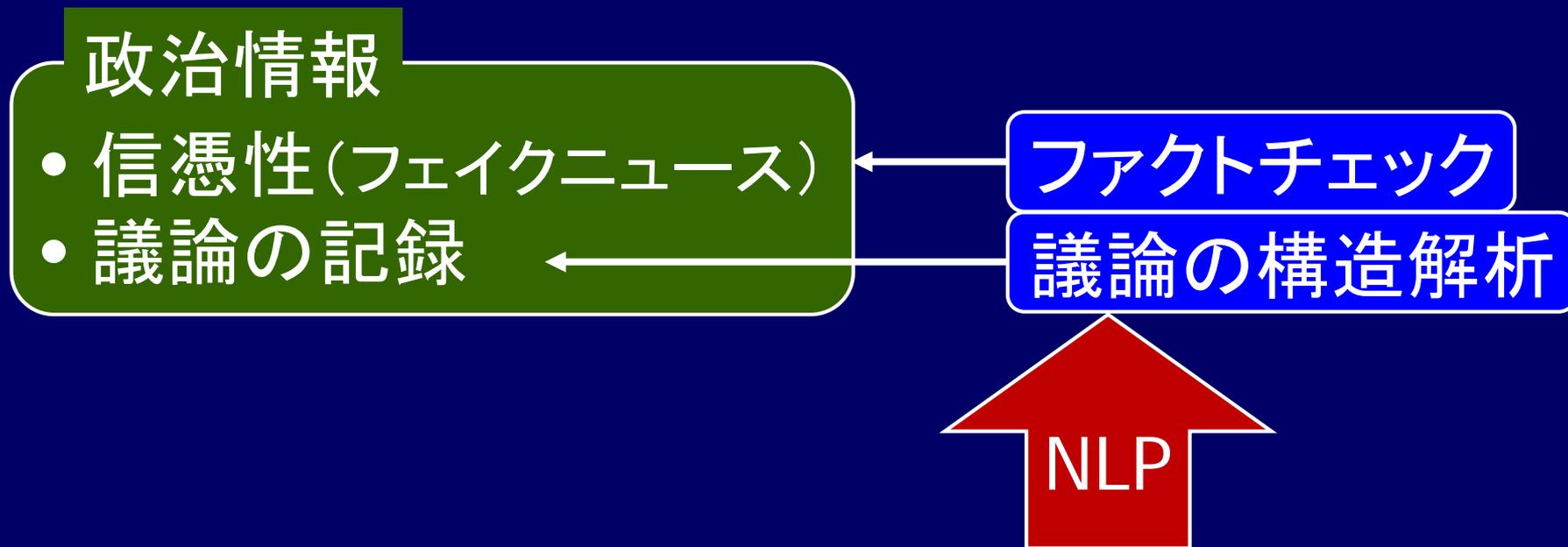
における

Dialog Summarization

小川泰弘 (名大),

木村泰知 (小樽商大), 渋谷英潔 (NII),
秋葉友良 (豊橋技科大), 横手健一 (日立),
乙武北斗 (福岡大), 石下円香 (NII)

背景 政治情報 + NLP



自然言語処理のアプローチにより
政治情報の課題を解決したい

NTCIR14 → NTCIR15

- NTCIR14 QA Lab-PoliInfo (2018-2019)
 - Segmentation task
 - Summarization task
 - Classification task
- NTCIR15 QA Lab-PoliInfo2 (2019-2020)
 - Stance Classification task
 - Dialog Summarization task
 - Entity Linking task

QA Lab-PoliInfo2 タスクイメージ

例えば、国会中継



Stance Classification

国民民主党

玉木雄一郎

自由民主党

安倍晋三

(理想はリアルタイム音声認識)

玉木雄一郎

「しかも、災害対応の陣頭指揮をとるべき石井国土交通大臣は、被災者救済よりもカジノ法案を優先するという対応をとり続けました。...

政府が今国会最大の目玉法案と提出した働き方改革法案も、国民の命をないがしろにしているとわざるを得ません。」

安倍晋三

「これまでの働き方改革の上に、生涯現役社会を目指し、六十五歳以上への継続雇用の引上げや中途採用、キャリア採用の拡大など雇用制度改革に向けた検討を進めます。」

Entity Linking

一次情報

会議録

平成xx年xx月xx日xxx議会

石井啓一

「この法律案は、…、世界最高水準のカジノ規制等によって、様々な懸念に万全の対策を講じるもの..(Evidence)」

表記ゆれ

Wikipedia

特定複合観光施設区域の整備の推進に関する法律

「.....」

働き方改革関連法

「.....」

独占禁止法

曖昧性解消

「.....」

私的独占の禁止及び公正取引の確保に関する法律

「.....」

Dialog Summarization

PoliInfo2

- Stance Classification
 - スタンス(会派)の分類
- Dialog Summarization
 - 議会の質問と答弁の構造を考慮した要約
- Entity Linking
 - メンション抽出
 - 曖昧性解消

DIALOG SUMMARIZATION

Dialog Summarization

- 目的
 - 地方議会会議録の対話構造を考慮した要約
- 入力
 - 東京都議会の会議録＋要約に必要な情報
- 出力
 - 「都議会だより」と同様の要約
- 評価
 - 参加者による人手評価
 - ROUGE scores

都議会会議録

東京都議会
Tokyo Metropolitan Assembly

文字サイズ **拡大** 標準

都議会の紹介 議員の紹介 会議の結果と記録 傍聴・見

トップ > 会議の結果と記録 > 本会議の会議録 > 平成30年第4回定例会

平成30年第4回定例会 会議録

日付	号数	内容
12月4日	第15号	議事・知事所信表明など
12月11日	第16号	代表質問 ・ 小山くにひこ（都ファースト） ・ 山崎一輝（自民党） ・ 橋正剛（公明党） ・ 尾崎あや子（日本共産党） ・ 宮瀬英治（立憲・民主）
12月12日	第17号	一般質問 ・ 石川良一（都ファースト） ・ たきぐち学（都ファースト） ・ 栗林のり子（公明党） ・ 清水孝治（自民党） ・ 斉藤まりこ（日本共産党） ・ 木下ふみこ（都ファースト） ・ 菅原直志（都ファースト） ・ 小林健二（公明党） ・ 伊藤しょうこう（自民党） ・ 森村隆行（都ファースト） ・ あかねがくぼかよ子（日本共産党） ・ ひぐちたかあき（都ファースト） ・ やながせ裕文（維新） ・ 議事

百十五番小山くにひこ君
（百十五番小山くにひこ君登壇）

○百十五番（小山くにひこ君） 東京都議会第四回定例会に当たり、都民ファーストの会東京都議団を代表して、小池知事及び教育長、関係局長に質問いたします。

いよいよ二〇二〇年の東京オリンピック・パラリンピック競技大会まで二年を切りました。一九六四年の東京大会は、戦後復興の象徴であり、首都高速道路や地下鉄の建設、東海道新幹線の開通など、各種インフラの整備が進みました。一九六四年大会後、日本は高度経済成長を続け、その後の日本と東京の発展へと大きくつながりました。

その後の平成は激動の時代でありました。バブル崩壊から始まった長期的な経済停滞、経済のグローバル化、IT化の流れの中で、日本の国際的地位は低下しました。一九六四年大会後に増加を続けていた日本の人口は、二〇〇八年をピークに既に減少に転じており、東京都の人口も二〇二五年をピークに減少に転じると見込まれております。このような社会経済情勢の劇的な変化は、戦後日本の成長を生んだ社会モデルからの変革を迫っております。

平成の時代が明年幕を閉じ、新たな時代を迎える成熟都市東京は、今まさに大きな変革を必要としています。少子高齢化による生産年齢人口が減少する中で、次なる成長の源泉となる人、物、金、情報をめぐる世界の都市間競争、まさに熾烈をきわめています。このような状況下において多様性が成長の源泉であると、そういった認識に立ち、二〇二〇年の東京大会とその先を見据え、世界の中で戦う東京の成長戦略を描き出さなければなりません。

そして、私たちは、一九六四年東京大会をきっかけに築き上げられてきた東京を二〇二〇年大会を契機として再構築し、東京と他の地域がともに栄える、東京の持続的成長を実現していかなければなりません。

私たち都民ファーストの会東京都議団は、都議会最大会派となり一年余が経過しました。この間、議会改革を初め、受動喫煙防止条例の制定、待機児童の大幅減少、オリンピック・パラリンピック憲章人権条例の成立など、二〇二〇年の先を見据えた東京の成長と発展の礎となる施策が着実に推進されてきました。

本定例会でも、中小企業の振興条例、防災対策、暑さ対策を柱とする補正予算など、未来の東京の成長と発展のために必要不可欠な施策が取り上げられております。

このような東京都の取り組みにもかかわらず、国はまた、不合理な都税の取奪を繰り返そうとしています。今、都議会に求められているのは、都議会一丸となって、他の地域との共存共栄を可能とする首都東京の成長戦略を描き出し、着実に実行することであると改めて申し上げ、以下質問いたします。

平成三十一年度税制改正について伺います。

国は、いわゆる偏在是正の名のもと、都の税財源を地方へと配分すべく、さまざまな措置を講じてきました。この間、都としても対抗策を講じてきましたが、平成に入ってからの三十年間で都が失った財源は六兆円に上り、平成三十一年度税制改正においても、さらなる措置が事実上予告されております。

こうした国の不合理な税制改正の動きに対して、先般、私たちの提案により立ち上げられました東京と日本の成長を考える検討会の報告書が取りまとめられ、また、東京都税制調査会の答申も示されました。そして、それらを受けた東京都の見解も示されて

一括質問一括回答

出の国会議員、与党税制調査会の国会議員、都内区市町村との折衝も、東京都選出の国会議員や与党税制調査会の国会議員への要請活

に提出をいたしました平成三十一年度予算の編成等に関する建議の

中で、平成財政の総括が明記をされております。

そこでは、少子高齢化に伴い増大する負担を将来世代に先送りしている負担先送りの罪深さ、債務残高対GDP比が第二次世界大戦末期の水準に匹敵をしている事実、税財政運営の大原則である受益と負担の乖離の拡大等に触れ、新たな時代において、財政健

都議会だより

- 都議会の広報紙
- 年5回発行
- Web上で閲覧可能
- 本会議の要約を掲載
 - サブピックごと
 - 質問-答弁の対応

東京都議会
Tokyo Metropolitan Assembly

サイトマップ モバイル English
文字サイズ 拡大 標準 キーワードを入力

都議会の紹介 議員の紹介 会議の結果と記録 傍聴・見学 調査・友好交流など

トップ > 都議会だより > 331号 > 代表質問 (小山くにひこ)

中小企業・小規模企業の支援を 幼児教育無償化への都の対応は

▶ 00:00 / 00:00

小山くにひこ (都ファースト)

産業振興

(1) 中小企業・小規模企業振興条例の理念に基づき、活力ある地域社会をつくり雇用の創出を。(2) 農業は東京の持続的成長に必要不可欠。農業振興への今後の展開は。

知事 (1) 地域経済の持続的発展と雇用創出の実現のため効果の高い振興策を展開。(2) 都市農地の保全、担い手の確保と育成・定着の体制整備、先進技術活用等、様々な施策を展開。

ダイバーシティ・東京

(1) 国の幼児教育無償化案では負担の軽減は十分とは言えず、また認可と認可外で格差が生じる。対応は。(2) 児童虐待対策の条例制定では未然防止の視点を重視して進めるべき。LINE相談の一層の活用も含め見解は。(3) 小中学校のスクール・サポート・スタッフの配置支援を拡大すべき。(4) 学校の働き方改革を加速させるため、部活動指導員をはじめ専門スタッフの質・量の確保を。(5) 受動喫煙防止条例の施行に向けて、内容の一層の周知徹底と実効性の確保を。

知事 (1) 待機児童対策協議会で国と意見交換。国の動きを踏まえ適切に対応。(2) 体罰等を行ってはならないこと等を未然防止の観点から条例に明記。LINE相談は31年度から本格実施。(5) 条例施行等のタイミングで効果的な広報を展開。都民や事業者の理解促進や機運の醸成を図り、受動喫煙防止の取組を進める。

教育長 (3) 区市町村教育委員会と連携しながら配置拡充を検討。(4) スタッフの安定的確保や資質向上をはじめとする多様な取組を検討。

消防活動

東京2020大会に向けて、消防団の活動における暑さ対策を充実すべき。

消防總監 警戒活動等の役割を担うことから、酷暑でも活動しやすい被服等の整備に努める。

ラグビーワールドカップ

テストマッチの運営面の取組を大会にどのように繋げるのか。

オリパラ局長 本番を見据えた取組を行い運営ノウハウの蓄積に努めた。更なる検証を深めて運営計画に反映し、大会の成功に繋げる。

要約の入力

- 都議会議録
- 要約に必要な情報
 - 日付
 - 質問者名
 - メイントピック
 - サブトピック
 - 答弁者名
 - 要約文字数

The screenshot shows the Tokyo Metropolitan Assembly website. At the top, there is a navigation bar with the logo '東京都議会 Tokyo Metropolitan Assembly' and links for 'サイトマップ', 'モバイル', and 'English'. Below the navigation bar, there are menu items: '都議会の紹介', '議員の紹介', '会議の結果と記録', '傍聴・見学', and '調査・友好交流など'. The main content area displays a video player with the title '中小企業・小規模企業の支援を幼児教育無償化への都の対応は' and a progress bar showing '00:00 / 00:00'. Below the video player, the name '小山くにひこ (都ファースト)' is listed. The page contains several sections with redacted content (yellow boxes): '産業振興' (with a redacted '知事' field), 'ダイバーシティ・東京' (with redacted '知事' and '教育長' fields), '消防活動' (with a redacted '消防總監' field), and 'ラグビーワールドカップ' (with a redacted 'オリバラ局長' field).

詳細

The screenshot shows the Tokyo Metropolitan Assembly website with several callouts pointing to specific features:

- Main topic:** Points to the breadcrumb navigation: 議会だより > 331号 > 代表質問 (小山くにひこ)
- 号数(日付に変換):** Points to the main title: 中小企業・小規模企業の支援を幼児教育無償化への都の対応は
- Question Speaker:** Points to the speaker's name: 小山くにひこ (都ファースト)
- Sub topic:** Points to the category: 産業振興
- Question Summary:** Points to the question text: (1) 中小企業・小規模企業振興条例の理念に基づき、活力ある地域社会をつくり雇用の創出を。(2) 農業は東京の持続的成長に必要不可欠。農業振興への今後の展開は。
- Answer Speaker:** Points to the answer text: (1) 地域経済の持続的発展と雇用創出の実現のため効果の高い振興策を展開。(2) 都市農地の保全、担い手の確保と育成・定着の体制整備、先進技術活用等、様々な施策を展開。
- Answer Summary:** Points to the answer text (same as Answer Speaker).

データサンプル

会議録データとの対応
(一部の学習データのみ)

```
{ "AnswerEndingLine": [540],  
  "AnswerLength": [50],  
  "AnswerSpeaker": ["知事"],  
  "AnswerStartingLine": [533],  
  "AnswerSummary": ["新視点の対応加えた防災対応指針を11月に策定し  
                    防災力向上の道筋を早期に示す。"],  
  "Date": "2011-06-23",  
  "ID": "PoliInfo2-DialogSummarization-JA-Dry-Training-Segmented-00002",  
  "MainTopic": "東京の総合防災力を更に高めよ<br>環境に配慮した都市づくりを",  
  "Meeting": "平成23年第2回定例会",  
  "Prefecture": "東京都",  
  "QuestionEndingLine": 291,  
  "QuestionLength": 50,  
  "QuestionSpeaker": "山下太郎(民主党)",  
  "QuestionStartingLine": 284,  
  "QuestionSummary": "更に高める取組が必要。",  
  "SubTopic": "東京の総合防災力"  
},
```

データサイズ

参照データ	ファイルサイズ
東京都議会 会議録データ	42MB

データセット	出題数	ファイルサイズ
学習用データ 対応付けあり	438	414KB
学習用データ 対応付けなし	325	292KB
評価用データ	254	161KB

タスクオーガナイザーによる実践

タスクオーガナイザーによる実践

- 前処理
- 構造解析
 - PoliInfo における Segmentation task の成果
- 要約
 - PoliInfo における Summarization task の成果

既存手法の適用

前処理

- 発言者名のマッチ

都議会議録	都議会だより
石原慎太郎	知事
中嶋正宏	オリパラ局長

公開
予定

- 都議会議録を質問者ごとに分割
 - 一括質問一括回答が全部含まれる

構造解析へ

構造解析

一括質問一括回答

→ 質問-答弁の組合せに再構成(対応付け)

＝要約範囲の特定

• サブピックと発言者名が手掛かり

➤ 発言者名が議会だよりに1回だけなら自明

➤ 複数回 名前が現れる場合(質問者や知事)

◇ 手掛かり句に基づいて段落に分割

手掛かり句を見つけるための正規表現

文献[1]より引用

pattern	regular expression
opening	^まず ^最初に ^初めに ^次に ^次いで ^最後に ^終わりに ^[-一二三四五六七八九十]+点目 ^[^、]+についてで(す あります ございます)(が けれど) ^終わり(ま で)す。 ^以上で ^ありがとうございます ^他の質問に(ついて つきまして)は
closing	伺い[^、]*ます。 お尋ね[^、]*します お答えください。 (見解 所見 答弁)を求め[^、]*ます。 (いかがで どうで)(しょうか すか)。 .+質問を(終わります 終了します)。

[1] Kanasaki K, Yong J, Kawamura S, et al. Cue-phrase-based text segmentation and optimal segment concatenation for the NTCIR-14 QA Lab-PoliInfo task. In NII Testbeds and Community for Information Access Research - 14th International Conference, NTCIR 2019, Tokyo, Japan, June 10-13, 2019, Revised Selected Papers, pp. 85-96, 2019.

構造解析

一括質問一括回答

→ 質問-答弁の組合せに再構成(対応付け)

= 要約範囲の特定

• サブピックと発言者名が手掛かり

➤ 発言者名が議会だよりに1回だけなら自明

➤ 複数回 名前が現れる場合(質問者や知事)

◇ 手掛かり句に基づいて段落に分割

◇ サブピックを含む段落を要約元の候補

→ なければ、ベクトル化して類似する段落を候補

◇ ヒューリスティックス(出現順が同じ)により決定

要約

- 抽出型要約

- ランダムフォレストの改良版
- 正例と負例の比率を変化させたモデル
 - ◇ 長さの上限まで順番に文を抽出

- 文短縮

- ルールベース
- 係り受け解析 + 重要度

モデルは文献[2]と同一
(学習データ未変更)

Progressive ensemble random forest

ランダムフォレストの各モデルにおける抽出数

Document ID	111	106	23	19	92
文数	45	11	34	8	13
 N N N N N P 5	0	0	0	0	1
 N N N N P 4	0	0	0	1	1
 N N N P 3	0	2	1	3	1
 学習データ 中の正例・ 負例の比率 N N P 2	0	5	2	3	3
 N P 1	1	9	3	7	5

各モデルを順番に適用して、長さの上限までの文を抽出

評価

- 人手
 - 参加者間で相互評価
 - 前回 PoliInfo の場合
 - ✧ content, formedness, total
- ROUGE

	recall							F-measure						
	N1	N2	N3	N4	L	SU4	W1.2	N1	N2	N3	N4	L	SU4	W1.2
表層形	0.404	0.130	0.076	0.046	0.337	0.168	0.160	0.307	0.098	0.058	0.035	0.253	0.123	0.159
原形	0.424	0.143	0.086	0.054	0.354	0.184	0.170	0.323	0.108	0.065	0.041	0.266	0.134	0.169
内容語	0.244	0.105	0.050	0.022	0.234	0.123	0.150	0.185	0.078	0.037	0.017	0.177	0.079	0.139

Leader Board で使用
(人手評価との相関が高い)

Leader Board

- 参加者の投稿結果を随時表示
 - ROUGEスコアを自動採点
 - 投稿者が許可すれば、結果をダウンロード可



成果の共有 (強制ではない)

- 最終結果
- 中間データ
 - 他人の対応付けデータの利用
- プログラム

より良い成果を

まとめ

- NTCIR15 QA Lab-PoliInfo-2 の紹介
 - Dialog Summarization task
- 参加者募集中
 - <https://poliinfo2.net/>
 - 要約だけの参加も可
 - End-to-End な手法も可
 - 参加登録締め切り 4月30日