# 評価に関して

## スケジュール

2020年 05月07日 Dry Run 開始 ※目的:参加者からエラーを報告してもらう

2019年 05月27日 New! 第2回オンライン説明会

2020年 06月30日 Dry Run 終了 ※目的:参加者からエラーを報告してもらう

2020年 07月01-12日 Leader Boardの一時停止期間(データの修正期間)

2019年 07月12日 タスク参加登録締切

2020年 07月13日 本テスト開始 2020年 07月31日 本テスト終了

#### 2020年 08月01日 - 07日 参加者による評価期間

2020年 08月08日 - 14日 評価結果の集計期間

2020年 08月15日 評価結果の参加者への返送

2020年 08月下旬第四回説明会(報告会)

2020年 09月01日 タスク概要論文一部公開

2020年 09月--日 NTCIR-15タスク参加者論文(ドラフト) 提出締切

2020年11月01日 NTCIR-15タスク参加者論文・タスク概要論文 提出最終締切

2020年12月9-11日 NTCIR-15 カンファレンス & EVIA 2020 (NII, 東京)

## 評価

- ◆ Leader Boardにおける評価指標(自動評価)
- ❖ タスクの目的達成における評価指標(人手評価)。

2020年 08月01日 - 07日 参加者による評価期間

## 各タスクの目的

Stance Classification

➤ 議会の発言から会派のスタンスを分類

**Dialog Summarization** 

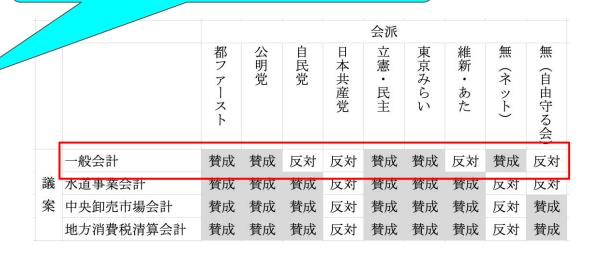
- ➤ <u>発言者の意図</u>と異ならないよう質問と答弁の構造を考慮した要約 Entity Linking
  - ➢ 発言中の表記を過不足なく抽出し、実体を提示する基礎技術

#### Stance Classification

#### 出力:2値と3値の2種類

```
"ProsConsPartyListBinary":{
 "都ファースト":"賛成".
 "公明党":"賛成",
 "立憲•民主":"賛成",
 "東京みらい":"賛成",
 "無(ネット)":"賛成",
 "自民党":"反対",
 "日本共産党":"反対",
 "維新・あた":"反対".
 "無(自由守る会)":"反対"
"ProsConsPartyListTernary":{
 "都ファースト":"賛成".
 "公明党":"賛成",
 "立憲・民主":"賛成",
 "東京みらい":"言及なし".
 "無(ネット)":"言及なし",
 "自民党":"反対",
 "日本共産党":"反対",
 "維新・あた":"反対",
 "無(自由守る会)":"言及なし"
```

#### 2値 ⇒ 自動評価:議案ごとの正解率の平均



3値 ⇒ 人手評価:「言及なし」が正解かを判定

#### Stance Classification

➤ 議会の発言から会派のスタンスを分類

# **Dialog Summarization**

#### 自動評価

- ROUGE-1(「内容語」の再現率)
- 制限字数を超えたものはO点

#### 人手評価

- Quality Questions (Well-Formedness, Content, Total)
- 3段階評価(Contentに関しては、「参照要約と異なるが適切」を含む4段階)

### **Dialog Summarization**

➢ <u>発言者の意図</u>と異ならないよう質問と答弁の構造を考慮した要約

# **Entity Linking**

正解:

以下を全て満たすもの

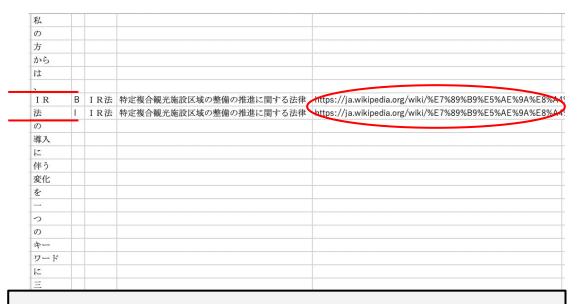
- ❖ 先頭の位置
- ❖ 末尾の位置
- ❖ 実体の記事

自動評価:

RecallとPrecisionのF値

人手評価:

特になし



## **Entity Linking**

➢ 発言中の表記を過不足なく抽出し、実体を提示する基礎技術

これを踏まえて、正解データは人手で作成 誤りや疑問はDry Run(~6月30日)中に報告