命題ネット構築に向けて

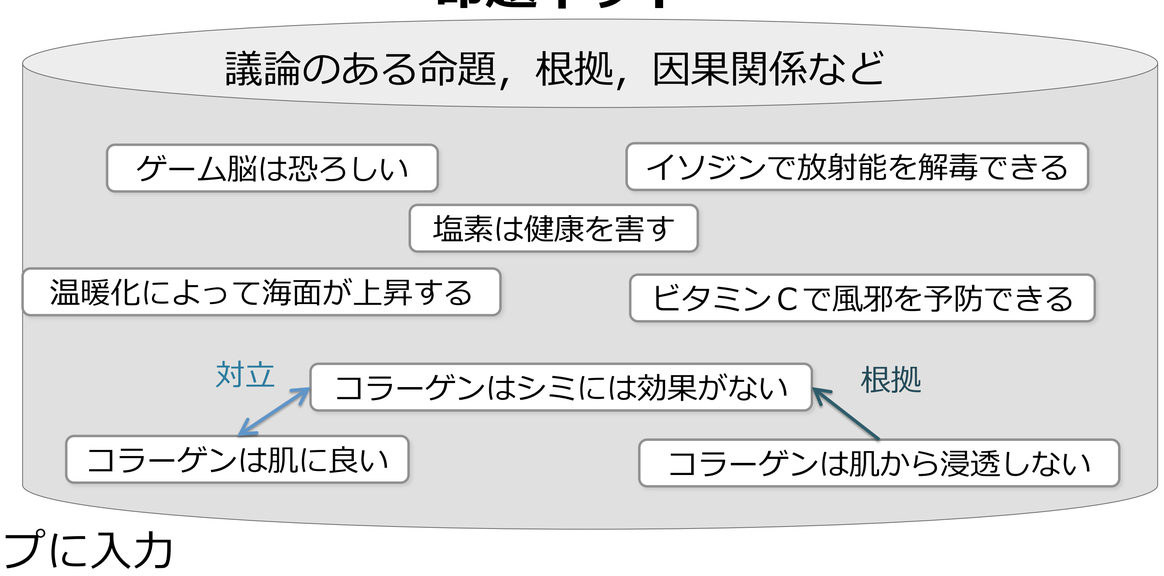
鍋島啓太, 水野淳太, 渡邉陽太郎, 岡崎直観, 乾健太郎(東北大学)

1.命題ネットとは

背景：個人の情報発信が容易になったことにより、流通する情報量が増えてきた。その一方、**真偽の不確かな情報も混在**。
真偽が不確かなものには、議論が存在すると思われる。
→議論を構成する命題間の対立関係、根拠関係を俯瞰したい

言論マップ[水野+,11]
- 入力されたクエリ文の関連情報を賛否に基づいて分類
- クエリ文に対して議論が存在するかは、言論マップが生成されるまで分からない（オンライン処理）

命題ネット
- 議論の存在する命題をWeb上から自動的に抽出し命題間の関係に基づいてグラフ化➔ オフライン処理
- 議論のより詳細な情報を得るには、命題をクエリとして言論マップに入力



関連研究:
- 流言情報クラウド[宮部+,12]
- マイクロブログ上の流言に対するユー ザの態度の分類[藤川,11]
- RTのコメントを使ったデマ分類[吉川+,11]

まずは命題を抽出する

2.命題抽出

抽出対象の命題：真偽について議論のある文
例：被曝対策としてイソジンを飲む
イソジンを飲むと放射線を防げる

震災関連ツイートコーパス
(約2億ツイート)

反論表現リスト
デマ
不確実
誤り
事実はない
・・・

①人手で整理した反論表現が含まれるツイートを抽出

コスモ石油の爆発により有害物な雨がふるというのは誤り

イソジンを飲むと被曝予防になるってデマが出回っている

⋮ 反論ツイート集合

②各単語について、その出現が反論ツイートに含まれる条件付確率上位1000単語を命題トピックとする

コスモ石油：0.5
築地市場:0.4
イソジン：0.2
雨：0.05
・・・

③命題トピックごとに、それを含むツイート中の文を抽出

コスモ石油 | 築地市場 | イソジン | ・・・

コスモ石油の爆発により有害な雨が降るというのは誤り

コスモ石油の件はデマ

⋮ 命題候補

④同じ命題トピックを含む文のうち、以下のスコアが一番高い命題候補を抽出 T:命題トピック, Word:内容語, s:文
hist(l,T)命題トピックTで長さlの文が出現する頻度

$$Score(s,T) = \left(hist(|s|,T) / \sum_{s \in S(T)} hist(|s|,T) \right) \sum_{word \in Words(s)} PMI(T,word)$$

命題候補中に、同じ内容語数の命題が
- 多い→重みを上げる
- 少ない→重みを下げる

トピックに関する議論を記述する際に必要な内容語を抽出

命題
コスモ石油の爆発により有害な雨がふる
イソジンを飲むと被曝予防になる
⋮

3.命題抽出の結果

命題トピックごとに、スコアが一番高い命題候補を抽出

命題トピック	命題	反論表現
田尻智さん	ポケモン生みの親・田尻智さんが亡くなった	不確定情報
ジャッキー	ジャッキー・チェン死亡説	デマ
ワンピース作者	ワンピース作者の尾田栄一郎さんが15億円寄付	デマ
コスモ石油	市原市のコスモ石油千葉製油所LPGタンクの爆発により、千葉県、近隣圏に在住の方に有害物質が雨などといっしょに飛散する	チェーンメール
イソジン	ヨウ素剤が現在手に入らないからといって、イソジンを飲む	誤報
日本医師会	日本医師会が（福島第一原発の）50km以内には立ち入らないようにと勧告した	デマ
化学薬品	千葉の製油所、製鉄所の火災の影響で科学薬品の含まれた有害な雨が降る	事実はありません
サーバールーム	サーバールームで怪我してて動けない	デマ
犬吠埼	犬吠埼に風力発電機を立てれば東京電力の電気量を全てまかなえる	デマ
物資投下	米軍ヘリの物資投下に辻元が抗議	デマ
双葉病院	福島・双葉病院にて患者を置き去りにして職員が逃げた	誤報

4.今後の課題

- 類似命題のクラスタリング
  - 「海藻類を食べると放射能の予防になる」「昆布を食べると被曝に効果がある」
  - 表記ゆれの吸収
    - 「ジャッキー」「ジャッキー・チェン」
- 一般的な単語で構成された命題の抽出
  - 「埼玉の水道水は危険だ」
- 根拠命題、対立命題の抽出