

# Web 議論における BERT を用いた 関連情報推薦エージェント

内藤 勝太<sup>1</sup> 白松 俊<sup>1</sup>

<sup>1</sup>名古屋工業大学 大学院工学研究科情報工学専攻

## 1. はじめに

多様な市民が参加する Web 議論では、建設的な合意形成を促進するファシリテーションが非常に重要であり、ファシリテーションを行う人をファシリテータと呼ぶ。2013 年に行われた Web 議論システム COLLAGREE を用いて実施された大規模社会実験[1]では、266 人による議論が行われた。実験結果および、議論参加者にアンケートからは、ファシリテータの役割である「炎上のような不適切な状態の回避」、「議論の発散・集約への適切な進行」がなされていたといったファシリテータの有用性を示す結果を得られた。

本研究で考える Web 議論における課題として、議論中の内容について背景知識を持っていない参加者が発言しにくいといった場面があると考えられる。そこで、議論内容に関連した理解促進につながる情報を議論参加者に提示することで、議論内容の理解や意見創出に繋がり、ファシリテータの議論進行の役に立つのではないかと考えた。そこで本研究では、議論内容についての情報を Web ページ検索することで収集、ページ中のテキストから情報を抽出し、議論参加者に関連情報を推薦するエージェントの開発を目指す。また、議論参加者への関連情報を読む労力の軽減を目指す。そこで以下の二つの課題が挙げられる。

- ① 議論内容から検索クエリを選び出し、情報を集める
- ② 収集した情報から推薦する情報を部分的に抽出する

一つ目の検索クエリを決める方法としてはシンプルな TF-IDF を用いて議論文脈から決定する手法と後続発言に含まれる単語の予測によるクエリ決定手法の二つを提案する。また、単語予測および二つ目の関連情報を抽出する際には、さまざまなタスクで高い精度を達成している言語表現モデルである BERT[3]を用いた手法を提案する。

### An Agent System for Recommending Relevant Information using BERT in Web-based Discussion

Shota NAITO, Shun SHIRSMATSU,  
Nagoya Institute of Technology

## 2. BERT

BERT(Bidirectional Encoder Representations from Transformers) [3]は、2018 年に Google が発表した自然言語処理のための汎用言語表現モデルである。Wikipedia などから得られる大量の文章データから事前学習しておいたモデルを、転移学習により文章理解や翻訳など様々なタスクに応用可能である。任意のタスクに対応させる学習のことを Fine-tuning, 転移学習と呼ぶ。本研究では、事前学習モデルとして日本語 Wikipedia を学習した公開モデルを使用した。

## 3. IBIS 構造



図 1 IBIS 構造

IBIS (Issue-Based Information System) [2]とはある問題に対する要素の相互関係をわかりやすく構造化したものである。IBIS 構造は大きく分けて以下の三つの要素に分類することができる。

- Issue : IBIS 構造のメインとなる要素。議論における問題点を質問形式で記述してある。
- Idea : Issue の解決に役立つとされる Issue への回答となる要素。
- Argument : 特定の Idea に対してメリット/デメリットおよびその他のステータスを持つ要素。

Issue の後にはその解決策となる Idea が続く。予測しやすいように、どの要素が先行文脈に含まれているかでその場に適した関連情報の推薦ができるのではないかと考えられる。関連情報としては、Merit(メリット), Demerit(デメリット), Solution(解決策), Example(事例), Reason(理由)の 5 つの種類を扱う。IBIS 構造の構成例を図 1 に示す。

## 4. 提案手法

### 4.1 クエリ決定手法

#### ① TFIDF によるクエリ

TFIDF による単語の重要度を表したスコアを求めたうえで、スコアの高いものをクエリとして使用する。議論内容の時間変化への対応をするため、発言が古いものに対して減衰率をかけた計算をすることで新しい発言に重点を置くことができる。また、減衰率をかけずに計算した結果もクエリに含めることで議論全体も考慮した検索ができるようにする。

#### ② BERT によるクエリ予測

関連情報には、検索したクエリと関連した単語が含まれているはずである。その単語の予測ができればクエリ決定ができると考えた。BERT を予測タスク用に学習させるために、YAHOO 知恵袋の質問回答データ [4] を利用し、文と単語のペアから成る学習データを作成し、転移学習により単語予測モデルの構築をした。モデルからの出力スコアの高い単語をクエリとして使用する。

### 4.2 関連情報判別手法

収集した情報の中から、議論文脈に沿った関連情報を選び出す必要がある。本稿では、BERT による関連情報判別手法を提案する。BERT に判別タスクを学習させるために、2013 年に Web 議論システム COLAGREE 上で行われた議論データに対しアノテーションしたデータを用いる。データ中から「idea-Argument」の関係を持つ文章のペアを抜き出す。Argument には 3 節で示した 5 種類の他に Answer と Opinion も含まれる。この 7 種と Other を加えた 8 種のラベルと文章ペアからなる学習データを作成し、関連情報判別モデルを構築した。本手法では Merit, Demerit, Solution, Example, Reason のスコアを利用し関連情報の選択を行っていく。

### 5. 関連情報推薦エージェント

本研究では、Slack 上での議論実験を想定している。関連情報推薦を行う際にはまず、Slack の API にリクエストすることによりその時点での議論データを受け取り、議論データを分析する。議論データから直近の参加者発言を取得し、TFIDF または BERT によるクエリ予測により関連情報検索に用いるクエリを決定する。次に、クエリでの検索を行い、検索結果から得られた Web ページからテキストを抜き出し、Web ページ毎で文のまとまり（本研究では 4 文でのまとまり）を作成する。本稿では、関連情報判別モデルによって文のまとまりを分類し、分類された数の最も多い関連情報について、スコアの高位上位三件を関連情報として推薦する。最後に、

Slack の API を用いて関連情報を投稿する。投稿する関連情報の中身は、関連情報の種類と Web ページタイトル、文章、URL の 4 つである。図 2 に議論実験中の Slack 画面を示す。



図 2 エージェントによる関連情報推薦

予備実験として、関連情報判別モデルで判別した文章について 5 種類それぞれの関連情報らしさをアンケートで評価した。その結果、訓練データの多い Solution と Example の評価が比較的高く、訓練データの少ない Demerit の評価が低い結果が得られた。

### 6. おわりに

本研究では、Web 議論システム上で動作する関連情報推薦エージェントの開発を行った。クエリ決定には TFIDF を用いた手法と BERT を用いた単語予測を用いた手法を提案した。また、関連情報判別には、BERT を用いた関連情報判別モデルを利用した手法を提案した。

今後の課題としては、二つのクエリ手法の比較し、BERT によるクエリ予測の評価をする。そして、関連情報エージェントを介入した議論実験を行い、被験者へのアンケートからエージェントの評価する。また、議論参加者からの意見を元にシステムの改善を行いたいと考えている。

### 謝辞

本研究の一部は JSPS 科研費 (17K00461) と JST CREST (JPMJCR15E1) の支援を受けた。

#### 参考文献

- [1] 伊美裕麻, 伊藤孝行ら “オンラインファシリテーション支援機構に基づく大規模意見集約システム COLLAGREE—名古屋市次期総合計画のための市民議論に向けた社会実装. 情報処理学会論文誌, 56(10), 1996-2010.
- [2] Noble, Douglas, and Horst WJ Rittel, “Issue-based information systems for design.” (1988).
- [3] Devlin, J. et al. “Bert: Pre-training of deep bidirectional transformers for language understanding.”, arXiv preprint arXiv:1810.04805.
- [4] ヤフー株式会社 (2019) “Yahoo! 知恵袋データ (第 3 版). 国立情報学研究所情報学研究データリポジトリ. (データセット).” <https://doi.org/10.32130/idr.1.3>