

Web 議論における BERT を用いた議論進行度の定量化手法

北川 晃¹ 白松 俊¹¹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

Web 上で何かの問題について議論をする時、行き詰まりが起こることがある。本研究では、議論の行き詰まりを検知したい。議論の行き詰まりを検知するためには、議論進行が停滞していることを検知する必要があるため、まず議論がどれくらい進行しているかということを定量化する必要がある。議論がどれくらい進行しているかの判断基準は、議論の状態によって異なると考えられる。そのため、議論の状態を発散フェーズと収束フェーズに分類した。本稿では、発散フェーズ・収束フェーズそれぞれの議論進行度の定量化について研究した。

2. 提案手法

人間の意思決定プロセスを記録するための手法の 1 つに、IBIS (Issue-Based Information System) [1] がある。これは、課題やアイデアなどをノードとして表現し構造化したものである (図 1)。本研究では、IBIS を用いて、議論進行度の定量化を行う。扱う IBIS のノードの種類は (現段階で) 課題・アイデア・メリット・デメリットに関する意見である。まず、ある時刻 t までの議論からテキストデータを読み込み、LSTM または BERT によって、IBIS のノードタイプをアノテーション付与していく。そして、各ノードのタイプに応じて重み (議論進行に対する貢献度) をつけ、全ノードの重みの和を時刻 t における議論進行度 $P(t)$ とする。傾き $(dP(t))/dt$ は重み (貢献度) の変化量であるため、議論の進行度を表していると考えられる。遺伝的アルゴリズム、及び被験者実験により得られる人間が感じる進行度のデータを用いることで、各ノードの最適な重み w を決定し、議論進行度の定量化を試みる。

3. データコレクション

各ノードタイプの重みを決定するために、過去に行われた議論を用いて被験者実験を行い、人間が感じる進行度をアンケートによって集めた。議論データとして、オンライン議論システ

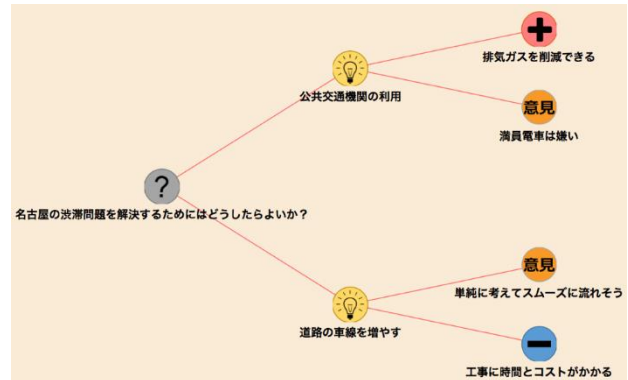


図 1 IBIS 構造

ム COLLAGREE を用いて 2013 年 11 月に行われた大規模社会実験のデータを用いる。被験者に議論内の各発言に関して、発散・収束フェーズ両方において、「発散/収束フェーズにおいて議論をどれだけ進展させたか」という観点で 0 から 5 の 6 段階で評価をしてもらった。上記の実験を発言数が 10 以上 21 以下の 17 個の議論それぞれに関して、13 人に行い、被験者それぞれ対して約 4 つの議論で実験をしてもらい、全 51 サンプルを集めた。

4. 評価実験

被験者実験によって得られた、各発言がどれだけ議論進行において重要となったかのアンケート数値の平均値を正解データとし、人間が感じる各発言の進行度とみなす。また、IBIS のノードに 0-1 の間の乱数で重みを設定することで、ノードの重みによって各発言の進行度を決定する。

遺伝的アルゴリズムによって、上記 2 つの進行度の相関が最大になるように、IBIS の各ノードの重さの最適化を行うことで、IBIS 構造のノードタイプの重みによる議論進行度の定量化を試みる。議論進行度の定量化には、下記の 2 種類の手法を用いた (図 2)。

手法 1. BERT と遺伝的アルゴリズムによる手法

BERT によって、テキストデータに IBIS のノードタイプをアノテーション付与することで、1 発言ごとに付与されたアノテーションに応じた議論進行度を決定する。

手法 2. BERT のみによる手法

アンケートによって得られた、人間が感じる 1 発言の進行度平均を正解データとし、BERT を用

d_1 : 手法1(BERT・遺伝的アルゴリズム手法)による議論進行度



d_2 : 手法2(BERT・テキストデータ手法)による議論進行度



図 2 手法 1, 2 の概要

いてテキストデータのみから議論進行度を予測する。

手法 1 と手法 2 で算出した議論進行度 d_1 , d_2 の重み付き線形和を用いて議論進行度 d を定義することで, 2 つの手法を相補的に用いる。

$$d = \alpha d_1 + (1 - \alpha) d_2 \quad (0 \leq \alpha \leq 1)$$

5. 結果

2013 年の COLLAGREE の議論を用いて各手法を試行した。その結果, アンケートによる人間が感じる議論進行度との相関は, 発散フェーズで最大+0.69, 収束フェーズで最大+0.44 となり, 人間の感じる進行度と中程度の正の相関を示した。各手法による相関を下記に示す(表 1)。

表 1 各手法の相関

手法	相関 (発散フェーズ)	相関 (収束フェーズ)
正解ノードタイプ (4/13種類) +GA	+0.44	-
手法1 BERTノードタイプ推定 (4種類) +GA	+0.47	+0.30
手法2 BERT(テキストのみ)	+0.62	+0.44
手法1・2による重み付き線形和	+0.69 ($\alpha=0.5$)	+0.42 ($\alpha=0.9$)

6. 考察

基本的には発散・収束フェーズ共に, IBIS のノードタイプの重さ最適化による進行度予測よりも, テキストデータからそのまま進行度予測の方が相関は高くなった。発散フェーズでは重み付き線形和を用いることで精度向上が見られたが, 収束フェーズでは精度向上が見られなかった。BERT による IBIS ノードタイプのアノテーション付与予測の精度がまだ低い(70 パーセントほど)ため, 有用なアノテーション予測ができていない可能性が考えられる。また, 人間が感じる議論進行度アンケートに関しても被験者間でばらつきが見られ(表2), より議論進行に対する知識を持っている人たち(実際のファシリテータなど)に協力してもらうことで精度向上が見込めるのではないかと考えられる。

表 2 被験者間の相関

アンケート被験者に関する相関平均	発散フェーズ	収束フェーズ
アンケート被験者同士の相関平均	+0.43	+0.39
正解データと1人の被験者の相関平均	+0.67	+0.74

7. おわりに

議論進行時に起こり得る行き詰まりを検知するために議論進行度の定量化を試みた。結果, 実用段階の精度ではないものの, 1 つの参考データとしてなり得る可能性があるのではないかと考えている。今後はこの進行度予測を用いて議論状態の予測を行なっていく。議論には 4 つの状態が存在すると考えられる。通常, 炎上, 停滞, 盛り上がりである。議論中の単位時間内の発言量と進行度の上昇量から, 以下のような関係性が考えられる(図 3)。これによって, 単位時間内の発言量と進行度の上昇量によって状態が変化する閾値を推定できたなら, 議論の状態を推測することが可能となり, よりファシリテーションの際に状態を考慮した発言を手助けできる可能性があると考えられる[2]。今後の展望としては, 状態変化の閾値の推定とともに, 進行度予測の精度上昇のために, アノテーション付与の精度上昇・人間が感じる議論進行度の精度上昇が課題となってくると考えている。

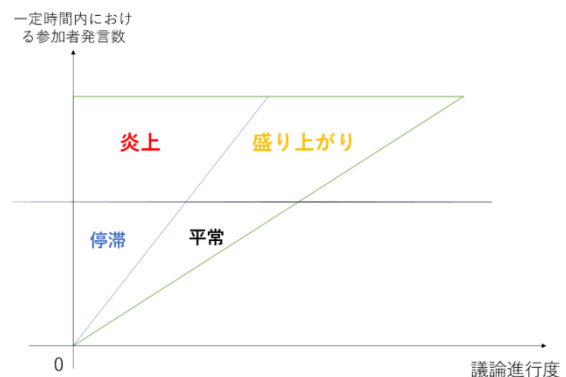


図 3 議論の 4 状態

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] Noble, Douglas, and Horst WJ Rittel. "Issue-based information systems for design." (1988).
 [2] Y. Ikeda, S. Shiramatsu. "Generating Questions Asked by Facilitator Agents Using Preceding Context in Web-based Discussion," In Proc. of IEEE ICA, pp. 127-132, 2017.