6E-01

DQI コーパスを用いた発言の自動分類と特徴表現の抽出

開發 大樹† 相馬 ゆめ‡ 大沼 進‡ 白松 俊†
名古屋工業大学† 北海道大学‡

1. はじめに

民主主義的な意思決定において，市民参加型の討議や熟議が重要視されている．しかし，どのような議論が望ましい意思決定に繋がるのかは明らかではない．討議中の発言の質を評価する指標として，DQI(Discourse Quality Index)が Steenbergen らによって提案されている[1]．しかし，従来の DQI では考慮されていなかった発言特徴（多元的な共通善を扱えない，議論の場にいない他者への配慮，当事者性の有無）があり，また異なる概念を１つの軸で評価している等の課題もあったため，相馬ら[2]はこれらを拡張した DQI を開発した．

本研究では，この拡張 DQI を付与した議論コーパス[2]を用いて，自動で議論の良さを判断するための指標ができないかを試行した．具体的には，同コーパスを用いて BERT[3]の分類モデルを再学習(fine-tuning)することにより，自動分類する実験を行った．さらに，分類の根拠を可視化する分析と，情報利得による特徴語抽出[4]を行った．

２．分析手法

対象データは，福島県内除去土壌の問題に関する議論を収録した，相馬らの DQI コーパス内の発言と特徴ラベルの組 576 件である．このコーパスは，議論中に話し合った内容が含まれる「本文」と，その文に付与される特徴ラベル２１種で構成されている．特徴ラベルは，「発言者の意見」「発言者の理由」「尊重なし」「尊重あり」「慮りなし（福島）」「慮りあり（福島）」「特定の地域・人々への言及」「社会全体の良さ（リスク・コスト・量）」**「社会全体の良さ（分かち合い・負担軽減）」「社会全体の良さ（福島の人々の気持ち）」**「疑問・論点」「対案・代替案」「まとめようとする発言」「妥協案」**「発言者の体験」「他人の体験」「当事者性」「リスク」「コスト・量」「風評被害・スティグマ」「発言者の感情」**で構成されており，太字で示したのが相馬らによって拡張された特徴ラベルである．特徴ラベルは１つの文章に対して２～３人の複数人で評価されており，一致率 k 係数は 0.61 であった．本研究で教師データとして利用する際，特徴ラベルは多数決方式で決定した．このデータに対し，BERT により発言への自動的に特徴ラベルを付ける学習を行う．また，情報利得によるラベルごとの特徴語句の抽出を行った．

３．BERT によるマルチラベル分類

議論テキストと各 21 個の特徴ラベルについて，BERT によるマルチラベル分類を行った．学習には，Simple-Transformers 内の Multilabel-Classification-Model ライブラリを用いて行う．

まず，訓練データとテストデータを固定して学習を行い精度の検証を行った．テストデータはランダムで抽出している．訓練データ数は 526 件，テストデータは 50 件とした．マルチラベル分類の指標の精度として LRAP(Label Ranking Average Precision)を用いる．
訓練データ 526 件テストデータ 50 件によるマルチラベル分類の学習における LRAP の値は 0.78 となった．

次に，モデルの汎化性を調べるために Cross-Validation(交差検証)を行った．交差検証とは学習データを N 個に分割し，n(n=0~N)番目の分割データをテストデータ，その他の N-1 個の分割データを訓練データとしてモデルに学習させ，精度を検証する．すべての分割データにおいて同様の操作を行う．これによって得られた N 個の検証結果の平均をモデルの汎化精度として扱う．本研究では 9 分割の Cross-Validation を行った．平均 LRAP の値は 0.836 となった．

表 1. 交差検証での Precision,Recall,F1-Score の値

特徴ラベル	P	R	F	件数
発言者の意見	0.94	1.00	0.97	543
発言者の理由	0.84	0.82	0.83	249
尊重あり(他の参加者)	0.83	0.55	0.66	244
コスト・量	0.70	0.45	0.52	108
社会全体の良さ(リスク・コスト・量)	0.67	0.43	0.49	92
社会全体の良さ(分かち合い・負担軽減)	0.49	0.31	0.34	46
社会全体の良さ(福島の人々の気持ち)	0.56	0.20	0.26	43
疑問・論点	0.39	0.16	0.21	182
慮りあり(福島)	0.42	0.09	0.14	60
リスク	0.19	0.03	0.05	90
対案・代替案	0.00	0.00	0.00	79
特定の地域・人々への言及	0.00	0.00	0.00	71
風評被害・スティグマ	0.00	0.00	0.00	43
まとめようとする発言	0.00	0.00	0.00	43

表 1 は，9 分割の Cross-Validation における各分割の Precision，Recall，F1-Score の平均を示している．F1-Score 上位10個と下位4個の結果を示す．

「発言者の意見」「発言者の理由」「尊重あり」「コスト・量」「社会全体の良さ(コスト・量)」の予測精度が F1-Score の観点において，他の特徴と比べて優れていた．理由としては，例えば，「尊重あり」の場合，「そうだと思います」などの文章が，「コスト・量」の場合，「～のコスト」など人間の目で見て明確に判別できる特徴だったということが挙げられる．

「尊重なし」「慮りなし」「妥協案」「発言者の体験」「他人の体験」「当事者性」「発言者の言及」「風評被害スティグマ」「リスク」はサンプル数が十分であるにも関わらず精度が良くなかった．理由としては，これらのラベルは人間にとっても基準がわかりにくいため人によって判断がわかれてしまうことが挙げられる．

次に，学習モデルがなぜそのラベルをつけたのかが客観的に理解できるように，Attention の可視化を行った．一例として,「尊重あり」と判定された文に関しての結果を示す．

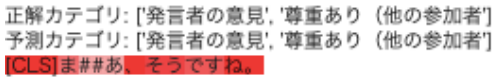

図 1.「尊重あり」の文章における Attention の可視化

図 1 は，「尊重あり」における Attention の可視化である．Attention が強くより判定に関わっているとされる語句は濃い赤で示されている．また，どの特徴がその語句に注目しているのかまでは実装していない．この語句が評価に影響しているということだけがわかる．「尊重あり」では，「そうですね」や「なるほど」など，同意を示すような文章が判断の基準であることがわかる．

図 2.「コスト・量」の文章における Attention の可視化

また，図 2 では「コスト・量」の文章での Attention の可視化である.「コスト」という語句が Attention により強く判定に関わっていることがわかる．

４．情報利得による特徴語抽出

情報利得と相互情報量を用いて特徴ごとに情報利得が高く相互情報量が正であるものを各特徴ラベルの特徴語句として抽出する．

表 2 は，特徴ラベルごとに情報利得が高いものから順に特徴語句を抽出した一覧である．

「尊重なし」からは「逆に」や「そうですけど」など他人の意見に逆説的になる語が抽出された．また，「尊重あり」からは「確かに」や「そうです」など他人の意見に肯定的な語が抽出された．

表2. 情報利得による各特徴ラベルにおける特徴語

語句	特徴	情報利得
思う_ので	発言者の理由	0.00058
風評_被害	風評被害・スティグマ	0.00427
安全_性	リスク	0.00154
逆_に	尊重なし	0.00143
確か_に	尊重あり	0.00181
健康_被害	リスク	0.00068
の_量	コスト・量	0.00064
そう_です	尊重あり	0.00156
怖い_な	発言者の感想	0.00096

５．まとめと今後の展望

本研究では，大沼らが拡張した DQI タグを付与した議論コーパスを用いて，BERT の分類モデルを学習することにより，自動で議論の良さを判断するための指標ができないかを試行した．DQI コーパスを用いて BRET による特徴ラベルの自動分類を実装し精度の検証を行った．また，Attention により分類根拠の可視化を行った．情報利得と相互情報量によって各ラベルで特徴的な語句の抽出を行った．

今後の展望としては，DQI コーパスを調整し，調整されたコーパスから再度学習を行い，精度の検証をする必要がある．

謝辞　本研究は，JST CREST(JPMJCR15E1)と NEDO(JPNP20006)の支援を受けた．

参考文献

[1] Macro R. Steenbergen, Andre Bachtiger, Markus Sporndli, Jurg Steiner. ``Measuring Political Deliberration: A Discourse Quality Index`` Comparative European Politics. pp. 21-48(2003).

[2] 相馬ゆめ，横山実紀，中澤高師，辰巳智行，大沼進. 公共的討議の「議論の質」の評価指標開発：低濃度除去土壌県外処理問題を題材とした集団討議実験. 日本リスク学会第 34 回年次大会講演論文集. pp. 116-121(2021).

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. ``BERT:Pre-training of Deep Bidirectional Transformers for Language Understanding. `` in Proc. NAACL-HLT 2019. pp. 4171-4186 (2019).

[4] Shun Shiramatsu, Takuya Nishida, Takayuki Ito, Katsuhide Fujita. ``Feature Expression Extraction from Discussion Facilitators'Utterances in Web-based Forum System towards Autonomous Facilitator Agents,'' in Proceedings of the 2016 5th IIAI International Congress on Advanced Applied Informatics, pp. 687--691 (2016).