

主題関連度を用いた課題達成のための対話の自動評価

Automatic Evaluation of Dialog for Task Achievement using Subject Relevance

萬徳 昭太 砂山 渡*
Shota Mantoku Wataru Sunayama

滋賀県立大学 工学部
School of Engineering, The University of Shiga Prefecture

Abstract:

対話の場面においては、時間の有益かつ効率的な利用が求められるが、対話を人手で評価することは容易ではなく、評価者の負担も大きい。そこで本研究では、課題達成のための対話の書き起こしデータについて、対話の中で用いられる主題に関係する単語との関連度を用いて、各発話者の主題に関係する発話量をもとに、対話の自動評価を行うシステムを提案する。

1 はじめに

対話の場面においては、時間の有益かつ効率的な利用が求められる。特に何らかの結論を出す必要がある対話においては、結論に結びつくアイデアを並べる発散フェーズと、挙げられたアイデアを整理していく収束フェーズとがあり、それぞれにおいて、結論に結びつく効果的な対話が行われることが望まれる。このような対話の場面においては、目的に沿った対話が行われているか、また対話に参加している全員が効果的な発言をしているかなどを、対話のリーダーシップを取る人が調整する必要がある。しかし、対話を人が客観的に評価することは容易ではなく、また評価者の負担も大きい。

そこで本研究では、課題達成のための対話の書き起こしデータについて、対話の中で用いられる主題に関係する単語との関連度を用いて、各発話者の主題に関係する発話量をもとに、対話の自動評価を行うシステムを提案する。これにより、対話の評価の自動化、ならびに対話の内容や方法の改善について、事後評価やリアルタイム評価などへの応用が期待できる。

2 関連研究

対話の場において、参加者の発言の自動評価を行い、音響的特徴と言語的特徴に基づく評価の分析結果をユーザにフィードバックをする研究 [1] や、前後の発言と兼ね合いを考慮しながら、時系列データとして発言内容

の説得力を評価する研究 [2] がある。本研究においては、対話の主題となる単語を設定した上で、発話に相当する対話中の各文の、主題との関連度をもとに対話全体を評価する。

議論能力に基づく Wikipedia における編集者間議論ページの分析を行う研究 [3] もあり、Wikipedia で行われている記事の修正や編集方針について、建設的な議論が行われることもあるが、編集者間の相互誹謗により無意味な論争が長く続く非効率な場合も見られる。そのため、建設的な議論の総量として、議論の主題と関わりをもとに対話の評価することが望まれる。

議論マップと呼ばれる議論を可視化するツールを作成する研究 [4] や、多人数が受講する講義における少人数で行う議論を対象として、他のグループの議論内容から活発さを反映した共起ネットワークを作成して提示する研究 [5] がある。本研究においては、対話の具体的な様子の理解ではなく、対話全体の総合的な評価を与える点数づけを行う。

3 主題関連度を用いた課題達成のための対話の自動評価システム

本章では、主題関連度を用いた課題達成のための対話の自動評価システムについて述べる。

3.1 本研究で用いる対話データ

本研究で用いる対話データは、指定の課題に対して新しい事業案を作成するために行われたもので、主に

*連絡先： 滋賀県立大学工学部 砂山 渡
〒 522-8533 滋賀県彦根市八坂町 2500
E-mail: sunayama.w@e.usp.ac.jp

表 1: 分析に用いた対話データと人手による評価点

グループ	人数	E1	E2	E3	評価点
A	4	○	○	—	8
B	4	◎	○	○	10
C	4	△	—	—	5
D	4	—	—	—	6
E	4	◎	—	—	8
F	4	△	—	△	4
G	4	◎	—	—	8
H	4	◎	○	—	9
I	5	△	△	△	3
J	4	◎	○	○	10
K	4	△	△	—	4
L	4	◎	○	○	10
M	3	○	○	○	9
N	4	○	○	○	9
O	4	—	—	△	5
P	4	—	—	—	6
Q	3	○	○	○	9
R	4	—	—	—	6
S	4	◎	—	○	9
T	4	○	○	—	8

4 名によって 20 分程度のオンライン対話が行われた音声データをテキストに書き起こしたものとなっている(表 1)。

本データは、ある人事関連サービスを行う会社から研究用にご提供いただいたもので、各対話データには、同社内で人手による、以下の評価が行われている。本研究では、その評価に基づいて、◎を 4 点、○を 3 点、—を 2 点、△を 1 点として、点数化した評価点を用いる。

- E1 : 総合評価
 - ◎大変優れている
 - ○優れている
 - —ふつう
 - △不足している
- E2 : 活動量
 - ○目立って多かった
 - —ふつう
 - △目立って少なかった
- E3 : 建設的議論
 - ○目立って見られた
 - —ふつう
 - △目立って少なかった

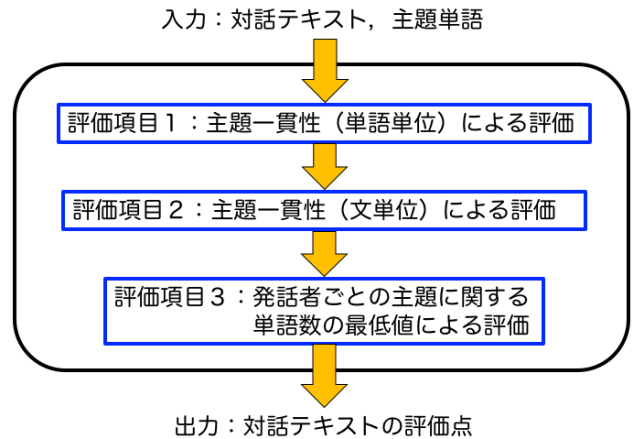


図 1: 主題関連度を用いた課題達成のための対話の自動評価システムの構成

3.2 主題関連度を用いた課題達成のための対話の自動評価システムの構成

主題関連度を用いた課題達成のための対話の自動評価システムの構成図を図 1 に示す。本システムには、評価対象となる対話テキストと、同テキストの主題を表す単語を入力する。システムは、入力された対話テキストに対して、入力された主題単語に関する 3 つの項目についてテキストの評価を行い、その評価結果を点数として出力する。

3.3 入力：主題単語の設定

本研究で用いる対話テキストに対して、設定した主題単語を表 2 に示す。これらの主題単語は、各対話テキストにおける頻度が高い名詞の中から、「対話のテーマ名に含まれる単語」または、対話の内容と関連のある単語として設定した。

3.4 評価項目 1：主題一貫性（単語単位）による評価

主題一貫性評価（単語単位）による評価値を、式 (1) で与える。ただし、式中の主題に係する単語は、文献 [6] による主題関連語としており、この値はフリーソフトのテキスト分析ツール TETDM[7] によって取得する。

$$\text{主題一貫性評価（単語）} = \frac{\text{主題単語または主題関連語の数}}{\text{単語の総数}} \quad (1)$$

表 2: 対話テキストに設定した主題単語

グループ	主題単語
A	盆栽 他 2 語
B	工芸品 他 2 語
C	子ども 他 1 語
D	生徒 他 1 語
E	オンライン 他 3 語
F	環境負荷 他 1 語
G	廃棄 他 5 語
H	インドネシア 他 4 語
I	インターシップ 他 1 語
J	教育 他 1 語
K	健康 他 1 語
L	中国 他 5 語
M	観光 他 3 語
N	人材 他 4 語
O	ボランティア 他 3 語
P	産業 他 3 語
Q	オンライン 他 4 語
R	能力 他 4 語
S	会社 他 3 語
T	仕事 他 3 語

表 3: 主題一貫性評価（単語単位）の評価点

評価点	主題一貫性（単語）[%]
4	40 % 以上
3	30 % 以上
2	20 % 以上
1	20 % 未満

主題関連語は、対話データの先頭の文から順に、主題単語または主題関連語が出現する文に出現する単語として逐次的に与えられる。また、主題一貫性評価（単語単位）の評価点を、表 3 のように与える。

本評価点を設定した理由は、課題達成に向けては、主題に関する単語を利用することが望まれる、と考えたことによる。

3.5 評価項目 2 : 主題一貫性（文単位）による評価

主題一貫性評価（文単位）による評価値を、式（2）で与える。ただし、式中の主題に関係する文は、文献 [8] による主題関連文としており、この値はフリーソフトのテキスト分析ツール TETDM によって取得する。

$$\text{主題一貫性評価（文）} = \frac{\text{主題に関係する文の数}}{\text{文の総数}} \quad (2)$$

表 4: 主題一貫性評価（文単位）の評価点

評価点	主題一貫性（文）[%]
4	20 % 以上
3	15 % 以上
2	10 % 以上
1	10 % 未満

表 5: 発話者ごとの主題に関する単語数の最低値の評価点

評価点	主題に関する単語数の最低値
4	15 個以上
3	12 個以上
2	10 個以上
1	10 個未満

主題関連文は、いずれかの主題単語または主題関連語が出現する文として与えられる。また主題関連語は、対話データの先頭の文から順に、主題単語または主題関連語が出現する文に出現する単語として逐次的に与えられる。また、主題一貫性評価（文単位）の評価点を、表 4 のように与える。

本評価点を設定した理由は、課題達成に向けては、基本的に主題に関する話をする事が望まれる、と考えたことによる。

3.6 評価項目 3 : 発話者ごとの主題に関する単語数の最低値による評価

本評価項目の値は、各参加者の主題単語または主題関連語を使用した回数を調べた上で、その参加者内での最低値として定義する。

主題関連語は、他の評価項目で定義した単語と同一のものをを用いる。また、主題に関する単語数の最低値による評価点を、表 5 のように与える。

本評価点を設定した理由は、課題達成に向けては、全員が主題に関する意見を述べる必要がある、と考えたことによる。

3.7 出力：対話テキストの評価点

前節までで述べた 3 つの評価項目による評価点を組み合わせる値を、対話テキストの評価点として出力する。表 6 に、本研究で用いた対話テキストの評価点を示す。

表 6: 各評価項目による対話テキストの評価点

グループ	評価項目 1	評価項目 2	評価項目 3
A	3	4	2
B	4	4	4
C	2	4	2
D	2	4	1
E	3	4	2
F	2	1	1
G	3	4	4
H	3	3	1
I	1	2	1
J	4	4	3
K	2	2	1
L	4	4	4
M	3	4	4
N	4	3	3
O	1	3	3
P	2	3	1
Q	3	4	4
R	2	1	1
S	3	4	3
T	4	4	3

4 対話の自動評価システムの評価

本章では、3 章の「主題一貫性（単語単位）」「主題一貫性（文単位）」「発話者ごとの主題に関する単語の数の最低値」の 3 つの評価項目の評価の妥当性とその組み合わせ方について検証した結果について述べる。

4.1 3 つの項目の組み合わせによる点数化の方法の妥当性の検討

表 7 に、人手による 3 つの評価項目による評価点数と、システムによる 3 つの評価項目による評価点数との相関係数を示す。

E1 から E3 の人による各評価項目について、システムの評価項目 1 の「主題一貫性（単語単位）」、または、評価項目 1× 評価項目 2+評価項目 3 が、およそ 0.8 以上の高い相関係数の値を示した。特に、人手による E1+E2+E3 の評価項目と、システムによる評価項目 1× 評価項目 2+評価項目 3 との相関が 0.90 と最も高い値となった。このことから、システムの評価によって、人手に近い評価が得られたことがわかる。

人手による評価項目ごとに見ると、全体的に評価項目 1 との相関が高い傾向にあった。また、E1 においては、評価項目 1× 評価項目 2+評価項目 3 との相関が最も高くなった。このことから、人手による評価においては、具体的な対話の内容に関わる評価では、単語単位での主題一貫性が重要視されていること、また対話

の結論に対しての総合評価としては、対話参加者全員の関与が必要な要素になっていると考えられる。

E1+E2+E3 の人手の評価の合計に対しては、評価項目 1 との相関と、評価項目 1× 評価項目 2+評価項目 3 との相関がほぼ同じ値となったが、先の考察の結果を踏まえると、3 つの評価項目を組み合わせ、評価項目 1× 評価項目 2+評価項目 3 によって評価していくことが望ましいと考えている。

5 おわりに

本研究では、課題達成のための対話の書き起こしデータについて、対話の中で用いられる主題に関係する単語との関連度を用いた 3 つの評価項目を提案し、対話の自動評価を行うシステムを提案した。提案する評価項目とその組み合わせにより、対話の自動評価を行える可能性を検証した。

今後は、対話の事後評価をもとにした、対話内容の改善を図るためのフィードバックを行うことや、リアルタイムの対話の中で、自動的に主題単語を設定した上で、リアルタイムに対話の評価を行うシステムなどへの応用を検討している。

謝辞

本研究の対話データは、(株) リクルートマネジメントソリューションズ組織行動研究所の今城研究員より、研究用データとしてご提供いただいたものです。ここに記して感謝申し上げます。

参考文献

- [1] 岡本康佑, 松原茂樹, 長尾確: 会議における発言の音響的特徴と言語的特徴に基づく自動評価, 情報処理学会第 78 回全国大会, pp.521 – 522 (2016)
- [2] 澤田慎太郎, 中川智皓, 新谷篤彦, 井之上直也: 対話的議論の自動評価に向けたディベートデータセットの構築, 言語処理学会第 26 回年次大会発表論文集, pp.708 – 711 (2020)
- [3] 朱成敏, 武田英明: 議論能力に基づく Wikipedia における編集者間議論ページの分析, 人工知能学会第 26 回全国大会, 1C2-R-5-1 (2012)
- [4] 趙雲超, 松村真宏, 谷内田正彦: 音声認識された議事録からの議論マップ自動生成, 人工知能学会第 20 回全国大会, 3C3-5 (2006)

表 7: 人とシステムの評価点の相関係数

評価 (人) / 評価 (システム)	1 (一貫性・単語)	2 (一貫性・文)	3 (主題最低発話)	1+2+3	1×2+3	1×3+2	2×3+1	1×2×3
E1 (総合評価)	0.78	0.64	0.62	0.78	0.82	0.75	0.72	0.74
E2 (活動量)	0.79	0.52	0.59	0.73	0.75	0.71	0.66	0.69
E3 (建設的議論)	0.78	0.57	0.64	0.77	0.78	0.79	0.73	0.77
E1+E2+E3	0.89	0.67	0.70	0.87	0.90	0.85	0.80	0.83

- [5] 村岡泰成, 石川誠彬, 尾澤重知, 江木啓訓: 議論の活発さを反映した共起ネットワーク図の提示による効果の検討, 情報処理学会第 110 回グループウェアとネットワークサービス研究会, No.8 (2020)
- [6] 砂山渡, 谷川信弘: テキストの主題に基づく一貫性評価と結論抽出への応用, 日本知能情報ファジィ学会誌, Vol.23, No.5, pp.727 – 738 (2011)
- [7] 砂山渡, 高間康史, 徳永秀和, 串間宗夫, 西村和則, 松下光範, 北村侑也: 統合環境 TETDM を用いた社会実践, 人工知能学会論文誌, Vol.32, No.1, NFC-A, pp.1 – 12 (2017)
- [8] 西原陽子, 佐藤圭太, 砂山渡: 光と影を用いたテキストのテーマ関連度の可視化, 人工知能学会論文誌, Vol.24, No.6, pp.480 – 488 (2009)