

授業グループワークにおける発話の意味量指標の提案とその効力の評価 Proposal and evaluation of a semantic quantity metric for utterances in group work

後藤 礼雄[†]
Reo Goto

市川 治[‡]
Osamu Ichikawa

1. はじめに

令和 3 年に文部科学省から発表された「学習指導要領の趣旨の実現に向けた参考資料」[1]において「協働的な学び」の重要性が指摘されてから、学習者同士の対話によって協働的な学びを目指した取り組みが増えている。しかし、グループワークのような複数の集団が同時進行する学習形態において、教師が各グループの状態や多数の学習者の対話を正確に把握、評価することは困難であり、各グループの状態を可視化することが求められている。

こうした背景を受け、近年はグループワークの対話場面から学びの過程を可視化し、評価を支援する手法が注目されている。ここでは「意味のある発話」、すなわち対話の流れを形成しうる発話がなされたかが、重要な評価視点となる。しかし、教師が児童生徒の発言記録から「意味のある発話」を抽出し評価することは多大な労力を要する。

そこで本研究では、対話中の発話の重要性を「発話の意味量」として定義し、客観的指標として提案することで、教師の評価支援を目指す。意味量の多い発話は、図 1 のようにその発話を契機に後続の発話が触発されてつながる（以降、係受け）ことが多く、さらに即時的な発話ほど影響を及ぼす確率が高いことが期待される。本研究の目的は、発話の意味量を定義、算出し、ある発話が後続の発話へ係受けする割合や、発話と時間間隔の関係を定量的に検証することである。

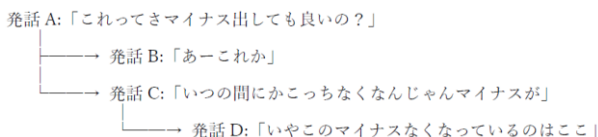


図 1 発話の係受けの例

2. 関連研究

これまで、対話データにおける「重要発話」の検出に関する研究が進んできた。対話を要約する際に重要なイベントに着目し、その発話を要約の特徴として利用する研究がある[2]。また、Nihei ら[3]は、会議の要約に寄与する発話を「重要発話」と定義し、会議中にそれらを推定するモデルを提案している。これらの先行研究では「要約に貢献する発言」を重要視するアプローチが多いが、本研究では「係り受け」という、議論を活性化させる起点となる発話を“重要”と捉え、その検出に取り組む点で差異がある。

3. データセット

使用するデータセットは、CoREF [2] から提供された実際の中学校のグループワーク授業を音声認識し、テキスト

[†] 滋賀大学大学院データサイエンス研究科 Graduate School of Data Science, Shiga University

[‡] 滋賀大学データサイエンス学部 Faculty of Data Science, Shiga University

	観点ありデータ	観点なしデータ
授業数	8	5
授業グループ数	86	118
行数	15306	13262

表 2 CoREF データの詳細

データとして書き起こしたデータである。今回使用したデータの詳細を表 2 に示す。本研究では、授業データの一部に人手で観点情報が付与されており、後述する目的に応じて観点ありデータと観点なしデータを使い分けている。本研究で特に重要なのは「後続発話への係受け」の有無だが、係受けラベルは観点なしデータにのみ付与されている。観点ありデータは、発話の意味量を算出する際に必要となる「不要発話」のクラスタベクトルを抽出するためにのみ使用する。

4. 提案手法

本手法は大きく 2 段階からなる。第 1 に、データ中の「不要発話」群を抽出するために、人手で付与された観点情報を用いてクラスタリングを行う。これにより、不要発話を取り除き、分析対象とする「主要発話」を明確化する。第 2 に、第 1 ステップで得られた「不要発話」クラスタベクトルを固定したうえで、対象データ全体に対して特殊なクラスタリングを実施する。そうして導出された「その他の発話」クラスタ以外のクラスタを「授業内トピック」クラスタとして考え、各発話と各トピックの中心ベクトルとのコサイン類似度を指標として与える。しかし、単純な受け答えや当たり前の発言は重要ではなく、非凡な発言、ありふれていない発言こそ重要であると考えたため、補助的に Perplexity を導入し、二つの指標の重み付き線形和として発話意味量を表現する。

以上により、「観点情報による不要発話群ベクトルの取得」→「トピックベクトルとの類似度+PPL による発話意味量の算出」という 2 段階構成で、発話意味量のスコアを算出する。

4.1 トピック関連度スコア

埋め込みモデル Sarashina-Embedding-v1-1B [3] を用いてベクトル化した発話 u のベクトルと、クラスタリングにより導出した各授業内トピックベクトルとのコサイン類似度を計算することで「トピック関連度スコア」を算出する。

4.1.1 トピックベクトルの導出

先述の通り、発話のトピック関連度を計算するためのトピックベクトルを、次の 2 段階 k-means クラスタリングにより抽出する。

まず、「観点ありデータ」を授業毎に分割、各授業の各発話埋め込み $e(u)$ に対し、クラスタ数 2 として k-means を実行する。そうして各授業で得られたクラスタのうち、いずれの観点とも高い重複を示さない「不要発話群」のクラスタ中心を、授業毎の独立クラスタベクトル $\mathbf{v}_{ind}^{(s)}$ として抽

出する。そして、全授業 S について抽出した $\{\mathbf{v}_{ind}^{(s)}\}_{s=1}^S$ を平均化し、最終的なグローバル不要発話クラスタベクトル \mathbf{v}_{ind} を得る。

次に、「観点なしデータ」を授業毎に分割し、クラスタ数をバイズ情報量基準 (BIC) で最適化した k-means を実行する。ただし、初期セントロイドの一つを前段で得たグローバル不要発話クラスタベクトル \mathbf{v}_{ind} に固定し、更新処理においても書き換えない。結果として得られる残りのクラスタ中心ベクトル $\{\mathbf{v}_k\}_{k=1}^K$ を、授業毎のトピックベクトルとみなす。

4.1.2 トピック関連度スコアの導出

前述で得られた各授業のトピックベクトル $\{\mathbf{v}_k\}$ とそのトピックに所属する発話 u の埋め込み $e(u)$ とのコサイン類似度を導出する。得られた全発話の各コサイン類似度 $\cos(e(u), \mathbf{v}_k)$ をトピック関連度スコアとして扱う。

4.2 Perplexity スコア

発話 u を語列 $w_{1:T}$ とする。n-gram 言語モデル \mathcal{M}_n を用いて発話の Perplexity (PPL) を計算する。今回は $n=3$ とした。n-gram の学習には、評価対象授業固有の語彙や言い回しに過剰適応することを避けつつ、評価対象のスコアを相対的に高くするために、評価対象以外の授業を用いた。また、スコアとして扱うため Perplexity に Min-Max 正規化を適用したものを最終的な Perplexity スコアとする。

4.3 発話意味量

トピック関連度スコア $S_{\cos-sim}(u)$ と Perplexity スコア $S_{ppl}(u)$ との両方を重み付き線形和として表現し、その値を発話意味量スコア $S_{meaning}(u)$ (以降、MS) として扱う、

$$S_{meaning}(u) = \lambda S_{\cos-sim}(u) + (1 - \lambda) S_{ppl}(u)$$

ただしパラメータ $\lambda \in [0, 1]$ は、コサイン類似度と Perplexity の影響度を制御する重み係数である。

5. 発話意味量の検証

「発話意味量」が後続への係受け及び遠隔発話へ与える影響を、次のような条件付き期待値を算出し、比較することで定量的に評価する

$$E(\text{短時間経過, 係受けする} \mid \text{高意味量}) \quad (1)$$

$$E(\text{短時間経過, 係受けする} \mid \text{低意味量}) \quad (2)$$

$$E(\text{長時間経過, 係受けする} \mid \text{高意味量}) \quad (3)$$

$$E(\text{長時間経過, 係受けする} \mid \text{低意味量}) \quad (4)$$

$$E(\text{係受けしない} \mid \text{高意味量}) \quad (5)$$

$$E(\text{係受けしない} \mid \text{低意味量}) \quad (6)$$

ここで、「意味量の高低」は、算出された MS の上位 25% を「高意味量」、それ以外を「低意味量」として二値化したものである。また、時間経過の短長は、後続発話との発話間隔 3 秒を閾値として設定した。

さらに、MS の算出にあたっては、重み λ を ① $\lambda = 1.0$ ($S_{\cos-sim}(u)$ のみ), ② $\lambda = 0.8$ ($S_{\cos-sim}(u)$ 重視), ③ $\lambda = 0.5$ (均等), ④ $\lambda = 0.2$ ($S_{ppl}(u)$ 重視), ⑤ $\lambda = 0$ ($S_{ppl}(u)$ のみ) の 5 通りで設定し、結果を比較した。各条件付き期待値を表 3 にまとめる。また、新たな指標として、我々の期待する仮説を立証するような値の一つのスカラ値として導入する。

$$\Delta I = ((1)-(2)) - ((3)-(4)) - ((5)-(6))$$

ΔI が大きいほど本研究の期待する仮説である「意味量の多い発話が後続発話を即時に誘発しやすい」に合致した振る舞いが現れていると判断できる。

λ	MS	短時間係受け	長時間係受け	係受けしない	ΔI
1	高	(1) 0.663 (2575/3882)	(3) 0.006 (25/3882)	(5) 0.375 (1453/3882)	0.162
	低	(2) 0.582 (6650/11424)	(4) 0.008 (91/11424)	(6) 0.454 (5183/11424)	
0.8	高	(1) 0.669 (2584/3860)	(3) 0.006 (25/3860)	(5) 0.371 (1433/3860)	0.174
	低	(2) 0.58 (6641/11446)	(4) 0.008 (92/11446)	(6) 0.455 (5203/11446)	
0.5	高	(1) 0.666 (2572/3860)	(3) 0.006 (24/3860)	(5) 0.373 (1438/3860)	0.168
	低	(2) 0.581 (6653/11446)	(4) 0.008 (92/11446)	(6) 0.451 (5198/11446)	
0.2	高	(1) 0.666 (2569/3860)	(3) 0.007 (26/3860)	(5) 0.373 (1438/3860)	0.167
	低	(2) 0.582 (6656/11446)	(4) 0.008 (90/11446)	(6) 0.454 (5198/11446)	
0	高	(1) 0.442 (1727/3904)	(3) 0.005 (20/3904)	(5) 0.575 (2246/3904)	-0.402
	低	(2) 0.658 (7498/11402)	(4) 0.008 (96/11402)	(6) 0.385 (4390/11402)	

表 3 各重みでの条件付き期待値, ΔI の比較

6. 結果と考察

$\lambda=0.8$ が最も ΔI 指標が大きくなった。高意味量発話の短時間係受け率は 0.669 (低意味量 0.580) で差分 +0.089、非接続率は 0.371 (低意味量 0.455) で差分 -0.084 となり、 $\Delta I = 0.174$ となった。一方、PPL 単独 ($\lambda=0.0$) では短時間係受けが高意味量 0.442 < 低意味量 0.658 となり $\Delta I = -0.403$ と仮説と逆転した。「意味量の高い発話は短時間で後続発話を触発しやすい」という期待は、 \cos 類似度を主体に少量の PPL を組み合わせさせたモデルで最も強く裏付けられた。

7. おわりに

本研究では、教師の評価支援のために客観的指標として「発話意味量」を提案、その実効性の検証を行った。検証の結果、提案した発話意味量指標は短期的な係受けを誘発し、対話を活性化する力を備えていることが示された。今後は、閾値設定の最適化や、さらなるデータセットの拡張、教育現場での有用性を検討していく必要がある。

謝辞

教育環境デザイン研究所および国立教育政策研究所の白水始氏からはデータの供与と貴重な助言をいただいた。また、本研究は JSPS 科研費 B24H00171a の助成を受けた。

参考文献

- [1] 文部科学省初等中等教育局教育課程課, “個別最適な学びと協働的な学びの一体的な充実”, 文部科学省, (2021). (参照: 2025-05-19), https://www.mext.go.jp/a_menu/shotou/new-cs/senseioun/mext_01317.html.
- [2] 教育環境デザイン研究所, “教育環境デザイン研究所 / CoREF”, 教育環境デザイン研究所, (2021). (参照: 2025-05-19), <https://ni-coref.or.jp/>.
- [3] Elena Filatova and Vasileios Hatzivassiloglou, “Event-Based Extractive Summarization.” In Text Summarization Branches Out, pages 104–111, Barcelona, Spain. Association for Computational Linguistics, (2004).
- [4] Fumio Nihei, Yukiko I. Nakano, and Yutaka Takase, “Meeting extracts for discussion summarization based on multimodal nonverbal information.”, In Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16). Association for Computing Machinery, New York, NY, USA, 185–192 (2016).
- [5] SB Intuitions, “sarashina-Embedding-v1-1b”, Hugging Face, (参照: 2025-05-21), <https://huggingface.co/sbintuitions/sarashina-embedding-v1-1b>.