

# 生成対話モデルを用いた授業発話の影響度の可視化に関する研究

## Visualization of the Impact of Classroom Utterances Using Generative Dialogue Models

大西 朔永<sup>\*1</sup>, 椎名 広光<sup>\*2</sup>, 保森 智彦<sup>\*3</sup>

Sakuei ONISHI<sup>\*1</sup>, Hiromitsu SHIINA<sup>\*2</sup>, Tomohiko YASUMORI<sup>\*3</sup>

<sup>\*1</sup>岡山理科大学大学院 総合情報研究科 数理・環境システム専攻

<sup>\*1</sup>Graduate School of Informatics, Okayama University of Science

<sup>\*2</sup>岡山理科大学 情報理工学部

<sup>\*3</sup>岡山理科大学 教育学部

Email: i22ed08bf@ous.jp

あらまし：小学校等の授業において、教員が省察に使用できる時間は少なく、短時間での省察が必要になっており、システム化が望まれている。本研究では、小学校の算数の授業発話に対して、生成対話モデルに用いる Transformer から発話生成時に得られる Attention Weight と、対話モデルの発話生成で算出できる発話の生成確率の変化を用いて、発話の影響度分析と可視化を行っている。

キーワード：授業における対話分析, GVT, Attention Weight, XAI, 発話の影響度

### 1. はじめに

小学校の授業では、教員は児童の状況を見ながら授業を進めており、教員が一方向的に授業を行っていることはまれである。児童の場合は授業について自分の学びの状況や意見、感想を同級生間で発話することが多くあり、児童は対話をしながら授業を進めていると考えられる。教員と児童や児童間では、一種の対話が成り立っており、授業の理解を促したり示したりする発話や対話を自動的に分析することができれば、教員に対して多くのフィードバックが可能となる。特に、教員の発話による児童の発話への影響の可視化は、重要な要素の1つである。

本研究では、自然言語処理分野の生成対話モデルの1つである拡張 GVTSC モデルを用いて小学校の授業対話を分析し、可視化を行っている。

### 2. 拡張 GVTSC

本研究では、ニューラルネットワークを用いた対話モデルの1つで我々が開発している事前にクラスタリングを用いて発話者の特徴を抽象化する拡張 GVTSC モデル<sup>(1)</sup>を用いて小学校の授業対話を分析し可視化を行っている。その概要図を図1に示す。

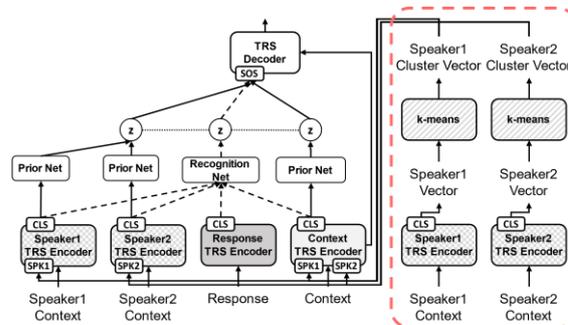


図1 拡張 GVTSC 対話モデル

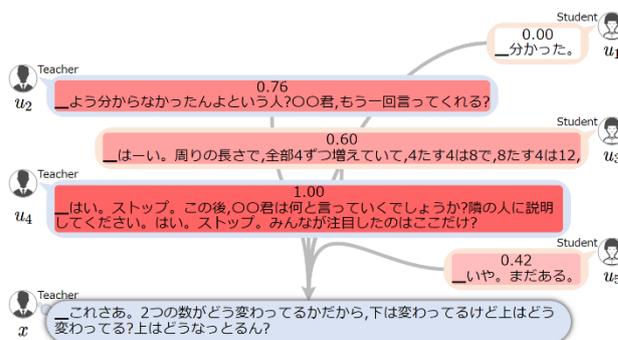


図2 Attention Weight による発話の影響度

### 3. 対話応答に対する発話影響の分析

自然言語処理に対する説明可能な AI(XAI, Explainable Artificial Intelligence)<sup>(2)</sup>の技術としては、Transformer 等の Attention を用いたモデルにおいて Attention の重みを可視化することで、予測の根拠とする手法が提案されている。また、入力に変化を与え、その出力の変化を定量的に分析することで、重要な入力部分を推測する手法が提案されている。その手法の一つとして、予測根拠としての妥当性を評価する ERASER ベンチマーク<sup>(3)</sup>が提案されている。本研究では、両手法を対話分析用に改良している。

#### 3.1 Attention Weight を用いた可視化

本研究では、対話モデルに拡張 GVTSC モデル内に Transformer を用いているため、Attention の重みを可視化することが可能である。対話応答生成において、ある対話応答を生成する場合、コンテキストの各トークンに対する重みをモデル内部で計算している。そのため、Attention の重みを利用することで、生成した応答にとって、重要なコンテキストのトークンの依存関係を分析することが可能であると考えられる。そこで、本研究では対話モデルの Attention

の重みを可視化することで、重要な発話の分析を行っている。Attentionの重みを可視化する手法は、  
(1) 応答の各トークンへの重みを平均し、コンテキストの各トークンへの重みも発話単位で平均する、  
(2) コンテキストの各発話への重みを0から1の値に正規化する、  
(3) 重みの大きさに応じて、発話を背景色で強調する。特に、モデルから得られるAttentionの重みは応答の各トークンとコンテキストの各トークン間の重みであり、コンテキストの各発話と応答の発話全体に対する影響度を測ることはできない。そこで、(1)の処理で発話全体に対する重みに変換している。

発話に対するAttentionの重みを可視化した例を図2に示す。図2の最後の発話は、応答発話であり、コンテキストから応答発話に向けて、影響を表す矢印の線を示している。なお、応答発話は強制的に実際の発話をモデルに生成させている。

### 3.2 対話応答の生成確率を用いた発話の影響度分析

ERASERベンチマークではAttentionの重み等の連続的なスコアを評価するComprehensivenessが提案されているが、対話における発話の影響度を分析するために改良した対話依存性指標(DD, Dialogue Dependency Index)を用いる。本研究では、対話モデルが授業発話のコンテキストから授業の応答発話を生成するとき、コンテキストから対象となる発話を除いた場合に、予測確率が低下するほど除去発話が重要であることを示す指標として対話依存性指標を用いる。対話応答の生成確率を用いた発話の影響度分析手法の概要を図3に示す。対話のコンテキスト  $c = u_1, u_2, \dots, u_N$  から発話  $u_i$  を除いた対話のコンテキスト  $c \setminus u_i$  から応答  $x$  が生成される確率  $p(x|c \setminus u_i)$  と、元の対話のコンテキスト  $c$  から応答  $x$  が生成される確率  $p(x|c)$  から対話依存性指標  $DD_i$  を  $DD_i = p(x|c) - p(x|c \setminus u_i)$  と定義している。教員の応答発話に対する児童の発話の影響度の可視化例を図4に示す。実際の教員と児童の対話の一部を例として、教員と児童の発話が最後の教員の応答発話にどの程度影響を与えたのかを可視化している。

### 3.3 分析手法の比較

Attention Weightを用いた手法と対話応答の生成確率を用いた手法を比較する。最初に、計算コストの観点では、Attention Weightを用いた手法は対話のコンテキストから応答を一度生成することで、Attention Weightを得ることができる。それに対して、対話応答の生成確率を用いた手法は、対話のコンテキストから順に発話を除いて応答を生成し、生成確率  $p(x|c \setminus u_i)$  を得る必要がある。元の生成確率  $p(x|c)$  も計算に必要なため、コンテキストの発話数を  $N$  とすると、生成回数は  $N + 1$  となる。計算コストの点ではAttention Weightを用いた手法の方が良い。

次に発話の影響度を求める精度の観点で比較を行

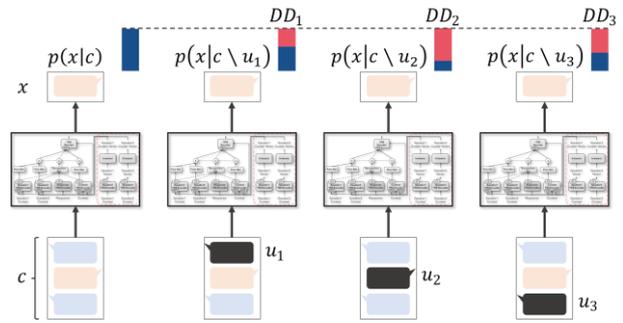


図4 対話応答の生成確率を用いた発話影響度

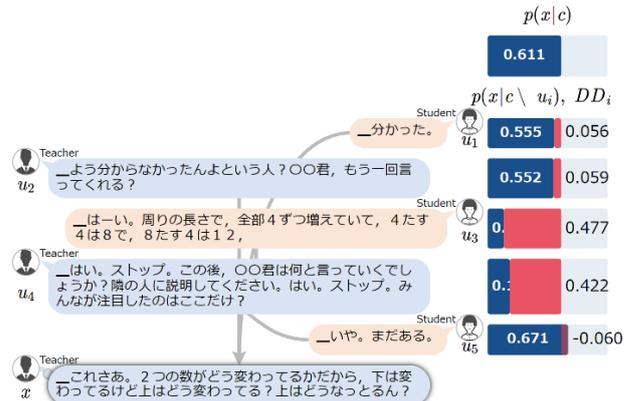


図5 発話影響度による可視化

う。両手法の分析例について、応答発話との関連が低い発話に対する評価を比較すると、Attention Weightを用いた手法は1発話目  $u_1$  のみが低い評価で、対話応答の生成確率を用いた手法は1, 2, 5発話目  $u_1, u_2, u_5$  が低い評価となっており、手法間で差がある。数の変化に関する応答発話との関連性では、1, 2, 5発話目  $u_1, u_2, u_5$  は低いため、対話応答の生成確率を用いた手法の方が高い精度であると考えられる。

## 4. 今後の課題

発話分析に用いる対話モデルの性能向上及び、授業データの収集、発話分析手法の改善が課題である。

### 謝辞

本研究の一部はJSPS 科研費 JP23K11378 の助成を受けたものです。

### 参考文献

- (1) 大西, 椎名, 保森: “対話モデルを用いた授業の発話分析とシーンの可視化”, 教育システム情報学会2022年度特集論文研究会, pp. 152-159 (2023)
- (2) Madsen, A., et. al.: “Post-Hoc Interpretability for Neural NLP: A Survey, Association for Computing Machinery, ACM Comput. Surv., Vol.55, No.8, pp.1-42 (2022)
- (3) DeYoung, J., et. al.: “ERASER: A Benchmark to Evaluate Rationalized NLP Models”, In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 4443-4458 (2020)