

## 総括的な振り返りのためのマルチモーダルエンベディングを活用した 学びの写真データのクラスタリングの試み

### A Study on Clustering Learning Photo Data by Using Multimodal Embeddings for Summative-Reflection

丸山 浩平<sup>\*1</sup>, 森本 康彦<sup>\*2</sup>

Kohei MARUYAMA<sup>\*1</sup>, Yasuhiko MORIMOTO<sup>\*2</sup>

<sup>\*1</sup> 東京学芸大学大学院連合学校教育学研究科

<sup>\*1</sup>The United Graduate School of Education, Tokyo Gakugei University

<sup>\*2</sup> 東京学芸大学 ICT/情報基盤センター

<sup>\*2</sup>ICT/Infrastructure Center, Tokyo Gakugei University

Email: r199003w@st.u-gakugei.ac.jp

**あらまし:** 一人一台端末による学習活動では、大切だと思ったことを写真に撮ることや撮影した写真を用いて学習活動に取り組むことが多い。写真を用いた学びの振り返りとして、学びの区切りの良い単元末や学期末に総括的に振り返ることで、自分のよい点や可能性、進歩の様子などを把握できると考えられるが、写真が多いほど総括的に振り返ることを難しくする。本研究では、写真を用いた総括的な振り返りの支援を目的とする。具体的には、写真データのマルチモーダルエンベディングを用いてクラスタリングし、類似した写真データのまとまりを児童生徒に見える化することで総括的な振り返りの支援を目指す。

**キーワード:** 総括的な振り返り、学びの写真データ、大規模視覚モデル、マルチモーダルエンベディング

#### 1. はじめに

児童生徒は、主体的に学習に取り組むことを通して、学習の見通しを立てたり学習したことを振り返ることが求められている<sup>(1)</sup>。近年では、一人一台端末を用いた学習活動の1つに、学んだことや大切だと思ったことを写真に撮ることや撮影した写真を用いて授業や学習活動に取り組むことが多く見られる。この中で蓄積される学びの写真は、学習活動のエビデンスとしてそのときの学びを想起しやすくし、これまでの写真から総括的に学びを振り返ることで、児童生徒自身のよい点や可能性などの多様な側面、進歩の様子などを把握しやすくすると期待される。

しかし、総括的な振り返りをする際、学びの写真が蓄積されるほど、その内容を確認することだけになったり、漠然と印象に残っている写真しか確認しなかったりする傾向がある。このため、これまでの学びの中でのよい点や可能性などの多様な側面、進歩の様子を振り返ることを難しくする。

一方近年、大規模言語モデル構築技術の進展により、テキストだけでなく画像の表す特徴の分析や画像の生成が注目され、画像の特徴を表したベクトル表現(マルチモーダルエンベディング)の取得と、それを用いたクラスタリングを行える利点がある。

そこで、本研究では、学びの写真データを用いた総括的な振り返りの支援を目的とする。具体的には、写真のマルチモーダルエンベディングを取得して、それをもとにクラスタリングし、類似した写真データのまとまりを構成する。そして、それを児童生徒に見える化することで、総括的な振り返りの支援を目指す。本論文では、学びの写真データのクラスタリングの試みについて述べる。

#### 2. 学びの写真データのクラスタリング

##### 2.1 画像のクラスタリング手法

画像データのクラスタリングの手法として、画像の色味に着目した手法と、画像の内容に着目した手法がある。前者は画像のRGB情報を抽出し、それをもとにクラスタリングするもので、画像の色合いが似たもののクラスタを構成できる。一方後者は学習済みのモデル(CNN, 転移学習など)に画像を読み込ませて情報を抽出し、それをもとにクラスタリングするもので、画像が表す内容に沿ったクラスタを構成できる可能性が明らかになっている<sup>(2)</sup>。

##### 2.2 マルチモーダルエンベディング

エンベディング(Embeddings)とは、テキストデータを高次元空間上における数値ベクトルに表したものであり<sup>(3)</sup>、主に大規模言語モデルの訓練過程で用いられている。マルチモーダルエンベディングとは、テキストだけでなく写真や動画などの種類に対応して表される数値ベクトルである。Google Vertex AI では大規模視覚モデル CoCa (Contrastive Captioner) をベースとした Multimodal Embedding Model に画像を読み込ませることで最大 1,408 次元からなるエンベディングを取得できる<sup>(4)</sup>。このベクトルデータをもとにクラスタリングすることで、似た内容の画像でまとまったクラスタを構成できる。

##### 2.3 学びの写真データのクラスタリングを活用した総括的な振り返り

総括的な振り返りとは、これまでに児童生徒が蓄積した多くのeポートフォリオ(学習データ)の中から最もよく取り組めたと思うことなどの意味あるeポートフォリオを吟味し精選することによって引

証づけるとともに、これまでの学習プロセスおよび学習成果やよい点や可能性などの多様な側面、進歩の様子を振り返るものである<sup>(6)</sup>。ここで、これまでに撮影した学びの写真データを確認する際に、類似した学びの写真でまとめられたクラスタを確認することで、これまでどのような学びに取り組んだのかを俯瞰して振り返る機会を与えられると期待される。そして、その中から自分がよく取り組めた学びの場面を吟味し精選しやすくできると期待される。その後、精選した写真について選んだ理由と今後の学びにどうつなげていくかを振り返って記録することで、総括的な振り返りを支援できると考えられる。

そこで、本研究では、学びの写真データのクラスタリングを活用した総括的な振り返りを提案する。

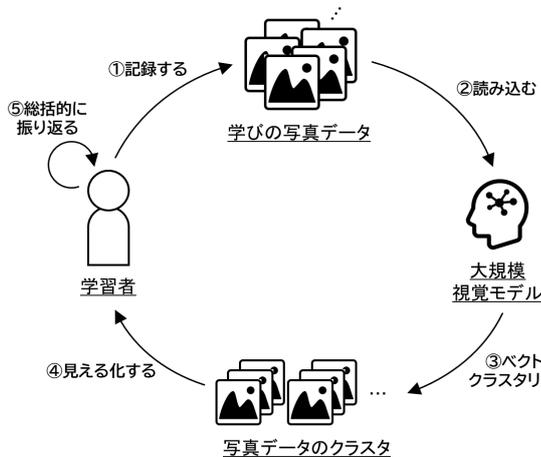


図1 学びの写真データのクラスタリングを活用した総括的な振り返りのイメージ

具体的には、学習者は学習活動の中で学んだことや大切だと思ったことを写真に撮り記録する（図1—①）。学びのある区切りにおいて、大規模視覚モデルに写真データを読み込ませ（図1—②）、得られたマルチモーダルエンベッディングを活用して写真データをクラスタリングする（図1—③）。その後、構成された写真データのクラスタを児童生徒に見える化し（図1—④）、児童生徒はそれを確認しながら総括的な学びの振り返りを行うものである（図1—⑤）。

### 3. 学びの写真データのクラスタリングの試行

#### 3.1 使用した学びの写真データ

マルチモーダルエンベッディングを活用した学びの写真データのクラスタリングにより、類似した写真データのまとまりを構成できるかを検証した。学びの写真データには、小学校の学習活動を想定した疑似データ9枚を用いた。

#### 3.2 マルチモーダルエンベッディングの取得

用意した写真データについて、Google Vertex AI Multimodal Embedding Model を用いマルチモーダルエンベッディングを取得した<sup>(4)</sup>。取得にあたってはPython クライアントライブラリを使用して写真データを読み込ませ、1,408次元のベクトルを取得した。

### 3.3 クラスタリング

まず取得したマルチモーダルエンベッディングについてt-SNEを用いて1,408次元から2次元に削減した。その後、エルボー法によるSSE（クラスタ内誤差平方和）の推移から4クラスタが妥当と判断し、k-means (scikit-learn)を用いてクラスタリングした。

### 3.4 結果と考察

用意した学びの写真データのクラスタリングの結果、4つのクラスタは表1のと通りの写真データ数で構成された。

表1 各クラスタに含まれる写真データ枚数

	クラスタ0	クラスタ1	クラスタ2	クラスタ3	計
枚数	1	3	3	2	9

実際の写真データと突合せてみると、クラスタ0は家庭科の被服実習の様子の写真、クラスタ1は黒板の板書やプリント、デジタルワークシートのスクリーンショットの写真、クラスタ2は体育科の体育館および校庭での練習の様子の写真、クラスタ3は算数の手書きしたノートの写真で構成されていることがわかった。

このことから、概ね類似した内容の写真データのまとまりでクラスタが構成されたことがわかった。このクラスタリングの結果を用いて、児童生徒が自身の学びの写真データを確認することで、自分がどのような学びや活動に取り組んできたのかを俯瞰して振り返ることができるようになることが期待される。そして、その中でも特によく取り組めたこと、印象に残ったことを吟味して精選しやすくなることが期待される。

## 4. おわりに

本論文では、総括的な振り返りのためのマルチモーダルエンベッディングを活用した学びの写真データのクラスタリングの試みについて述べた。今後は、学びの写真データの枚数を増やした検証と、最適なクラスタ数の検討、クラスタリングの結果を踏まえた児童生徒への見える化の仕組みの検討を行う。

### 謝辞

本研究は、科研費（23K02681）の助成を受けた。

### 参考文献

- (1) 文部科学省: “高等学校学習指導要領(平成30年告示)解説 総則編”. 東洋館出版社, 東京 (2019)
- (2) 柳澤秀彰, 山下拓朗, 渡辺裕: “主要キャラクタの抽出を目的とした漫画キャラクタ画像のクラスタリング”. 映像情報メディア学会誌, 73巻, 1号, pp.199-204 (2019)
- (3) OpenAI: “Embeddings”. <https://platform.openai.com/docs/guides/embeddings> (参照日: 2024.05.29)
- (4) Google: “Get multimodal embeddings”. <https://cloud.google.com/vertex-ai/generative-ai/docs/embeddings/get-multimodal-embeddings> (参照日: 2024.05.29)
- (5) 森本康彦ほか: “教育分野におけるeポートフォリオ”. ミネルヴァ書房, 京都 (2017)