

クラスタの“意味”を語る AI: LLM による教師なし学習の説明性付与

Pinheiro de Moura Caio César¹ 新見 佳祐¹ 山科 勇輔¹ 塩見 和真¹ 市川 佳彦¹
加茂 司² 久保谷 篤² 鮎澤 雄治² 山本 達也² 吉田 圭吾²
¹ 株式会社 Insight Edge ²SCSK 株式会社
{caio.pinheiro, keisuke.niimi, yusuke.yamashina,
kazuma.shiomi, yoshihiko.ichikawa}@insightedge.jp
{t.kamo, Atsushi.Kuboya, y.ayusawa, tats.yamamoto, keig.yoshida}@scsk.jp

概要

クラスタリング結果の解釈には、統計的知識に加えて対象分野のドメイン知識が不可欠であり、その多くは分析者の経験や主観に依存する。このため、解釈コストの増大や客観性の確保が課題となっている。本論文では、大規模言語モデル (Large Language Model; LLM) を用いてクラスタリング結果に説明性を付与する手法を提案する。クラスタごとの統計量とドメイン固有のコンテキスト情報を LLM に入力し、各クラスタの意味や特徴を自然言語で生成する。脳波 (Electroencephalography; EEG) データを用いた検証実験により、LLM が数値特徴と専門知識を結び付け、有用な解釈を提示できることを確認した。

1 はじめに

1.1 データ分析におけるクラスタリングの重要性

現代のデータ分析において、教師なし学習の一種であるクラスタリングは、ラベル付けされていない大規模データから潜在的な構造や特徴的パターンを探索的に抽出するための基盤的手法として重要な役割を果たしている。製造業における工程状態の分類、医療分野における患者群の層別化、マーケティングにおける顧客セグメンテーション、画像・信号処理分野における特徴分布の把握など、幅広い応用分野で利用されている。

代表的なクラスタリング手法である K-means 法は、計算効率の高さ、アルゴリズムの単純さ、および大規模データへの適用可能性といった利点を有しており、学術研究のみならず実務的分析においても

広く採用されている。そのため、K-means 法を中心としたクラスタリング結果の活用可能性を高めることは、実応用においても重要な課題となっている。

1.2 課題：解釈の高コストと客観性の欠如

一方で、多くのクラスタリング手法は、数値特徴空間における距離や類似度のみに基づいてデータを分割するため、得られたクラスタがどのような意味的特徴を持つのかを人間が直感的に理解することは容易ではない。その結果、クラスタリング後の分析においては、各クラスタに対して専門家が個別に特徴量分布や代表サンプルを確認し、意味付けを行う必要があり、解釈には多大な人的コストと専門知識が要求される。さらに、クラスタ数の設定や解釈の過程は分析者の経験や主観に依存する場合が多く、同一データに対しても解釈結果が分析者ごとに異なるという問題が生じる。このような主観性は、結果の再現性や客観性を損なう要因となり、特に意思決定支援や説明責任が求められる応用分野においては大きな制約となる。

1.3 LLM の台頭と構造化データ理解

こうした課題に対し、近年の LLM の発展は、新たな解決策の可能性を提示している。近年の LLM は、数値や統計量を含む構造化データを入力として受け取り、高度な推論、要約、およびデータの比較を行う能力を示している。この能力を活用することで、数値的なクラスタリング結果を神経生理学的な自然言語表現へと直接変換できる可能性が生まれている。

1.4 提案手法：統計量と専門知を統合する説明性付加

本論文では、LLM を用いてデータの特徴量とドメイン固有の専門知識を統合することで、クラスタリング結果に高度な「説明性」を付与する新しい枠組みを提案する。

本手法の有効性を検証するため、本研究では、信号の多次元性や個人差が大きく、解釈に高度な専門知識を要する EEG データを対象とした解析を行う。EEG 解析は、従来のクラスタリング手法において結果の解釈が特に困難な領域であり、本手法の説明性向上効果を評価するための適切な検証対象である。

- **手法の概要:** クラスタ重心（スケーリング前およびスケーリング後）を LLM に入力し、統計的特徴と専門知を結びつけた意味付けを自動で行わせる。
- **検証対象:** 本手法の有効性を検証するため、解釈が非常に複雑で専門性が求められる EEG を対象とした解析を実施した。

1.5 XAI の新たな手法としての位置づけ

本研究は、従来の XAI (Explainable AI) が主に教師あり学習における予測モデルの判断根拠を明示することを目的としてきたのに対し、教師なし学習に内在する「意味の空白」を埋めることを目的とした新しいアプローチを提示する。本手法は、数値空間における統計的構造と人間が理解可能な概念表現とを橋渡しする枠組みであり、解釈性が求められる多様な教師なし学習タスクにおいて、有効な説明可能性を提供する。

2 関連研究

2.1 EEG におけるクラスタリング研究

EEG データに対するクラスタリングは、潜在的な脳状態や神経生理学的パターンを探索的に同定する手法として広く用いられてきた。

特に、教師なし学習に基づくクラスタリングは、事前のラベル情報を必要とせず、EEG の周波数特性や頭皮分布に内在する構造を明らかにできる点で有用である。

代表的な手法としては、計算効率と実装の容易さから K-means 法が多用されており [1, 2], 周波数帯域別パワーや電極ごとの特徴量を用いた被験者単位

または時間窓単位のクラスタリングが報告されている。しかし、これらの研究においては、得られたクラスタの意味付けは分析者の経験や既存知識に依存する 경우가多く、解釈の客観性や再現性が課題として残されている。

2.2 LLM による説明性付加

近年、人工知能 (AI) による解析結果に対して、その判断根拠や意思決定過程を人間が理解可能な形で提示する説明可能性が強く求められている。

ニューラルネットワークをはじめとする機械学習モデルの出力は、しばしば数値的・抽象的であり、その不透明性が実運用や意思決定の妨げとなることが指摘されている。

製造業分野では、異常検知結果に対して LLM を用いた説明生成の試みが報告されている [3] 一方、医療分野においては、モデル出力の解釈が困難であることから、臨床応用に対して慎重な姿勢がとられる傾向がある。特に、教師なし学習に基づく解析結果は、予測対象や正解ラベルが存在しないため、説明の枠組み自体が十分に確立されていない。

2.3 本研究の位置づけ

EEG に基づいて生成された報告書やテキスト情報をクラスタリングし、それらを LLM に入力して要約や整理を行う研究は報告されている [4]。

しかし、数値特徴量に基づいて得られた教師なし EEG クラスタそのものに対して、LLM を用いて意味付けや説明を付与する試みはほとんど行われていない。

本研究は、EEG の統計的特徴量とクラスタ構造を明示的に LLM へ入力し、専門知識と結び付けた説明を自動生成する点に特徴があり、従来の EEG クラスタリング研究および XAI 研究の双方を補完する新しい枠組みを提供するものである。

3 提案手法

3.1 データセット

本研究で用いたデータセットは、OpenNeuro において公開されている ds004504 である [5]。本データセットは、アルツハイマー病患者、前頭側頭型認知症患者、および健常対照群の全 88 人分から構成されており、認知機能の異なる被験者群の脳活動パターンを比較解析するのに適している。

被験者は、複数の認知課題から構成される MMSE (Mini-Mental State Examination) を施行中に EEG 計測を受けており、各課題への応答過程における脳波が連続的に記録されている。EEG は 19 チャンネル構成で、サンプリング周波数 500 Hz で記録されており、各被験者あたりの計測時間はおおよそ 5 分から 15 分である。公開データには、アーチファクト (脳波以外のノイズ) 除去やフィルタリングなどの前処理があらかじめ施されている。

3.2 前処理

本研究では、各被験者から取得した EEG データ (図 1) を、長さ 5 秒、50% のオーバーラップを持つ時間ウィンドウ (図 2) に分割し、FFT 変換 (図 3) 行った後、delta, theta, alpha, beta, gamma の各周波数帯域に分解し、ウィンドウ単位で解析を行った。MMSE における最も容易な質問に回答するために必要と考えられる時間がおおよそ 5 秒であるとの仮定に基づき、5 秒の時間ウィンドウを採用した。

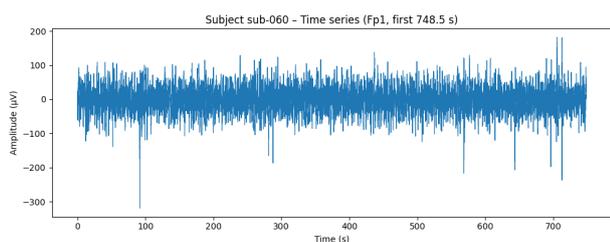


図 1 ウィンドウ分割前 (Fp1 チャンネル)

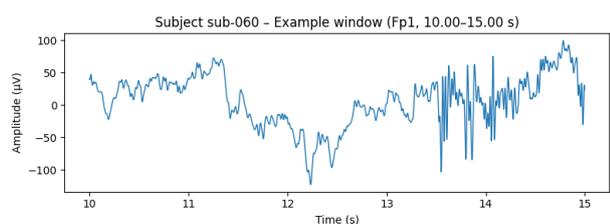


図 2 ウィンドウ分割後 (5 秒長)

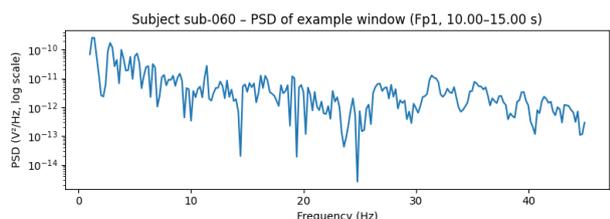


図 3 FFT 後

3.3 クラスタリング

3.3.1 手法

本研究では、クラスタリング手法として K-means 法を採用し、全 88 人分のデータに対して、ウィンドウ毎の脳波パターンのクラスタリングを行った。クラスタ数は、Elbow 法を利用し、 $k = 6$ に設定した。

3.3.2 クラスタリング結果

クラスタリング結果の 3 つのサンプルを以下に示す (図 4: アルツハイマー病患者, 図 5: 健常者, 図 6: 前頭側頭型認知症患者)。これらより被験者のクラス (アルツハイマー病患者, 前頭側頭型認知症患者, 健常者) 毎に現れる脳波パターンクラスが異なっていることが確認できる。

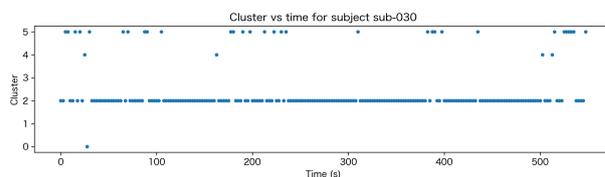


図 4 アルツハイマー病患者のクラスタリング結果

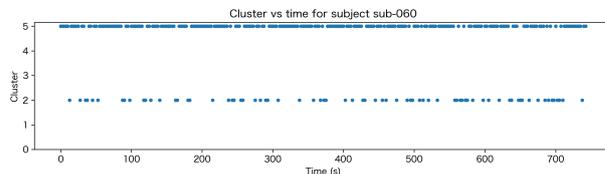


図 5 健常者のクラスタリング結果

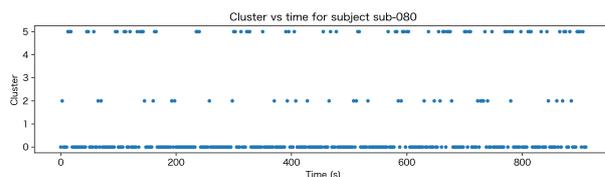


図 6 前頭側頭型認知症患者のクラスタリング結果

被験者クラス毎の脳波パターンの出現割合を図 7 に示す。これからも、健常者はクラスタ 0 の脳波パターンが多く、アルツハイマー病患者や前頭側頭型認知症患者ではクラスタ 5 の脳波パターンの出現割合が多くなっていることが確認できる。

3.4 プロンプト設計

前処理およびクラスタリングの完了後、各クラスタの特徴量を代表するクラスタ重心を LLM に入力し、統計的特徴と神経科学的・臨床的知見を結びつ

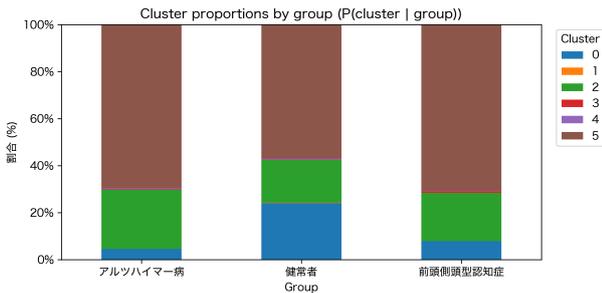


図7 sub-080

けた解釈を自動的に生成させた。この際、診断ラベルやその分布に関する情報は与えず、重心に含まれる特徴量のみに基づいて説明を行わせた。使用した LLM は Gemini 2.5 Pro である。

4 結果

LLM が生成した説明文を要約した結果を表 1 に示す。多くのクラスタにおいて、周波数帯域ごとの特徴に言及した説明が生成されており、既存の EEG 研究で報告されている神経生理学的知見と比較した結果、概ね整合的であることが確認された。特に、生理的脳波、病的脳波、およびアーチファクト由来と考えられるパターンが、説明文の内容から明確に区別されている点は注目に値する。

実際に、健常者では、図 7 に示すように、クラスタ 0 に割り当てられるウィンドウの割合が最も大きい。これに対応して、LLM はクラスタ 0 を「顕著なアルファ波の亢進を特徴とする、典型的な閉眼安静覚醒状態の脳波パターン」と説明しており、健常被験者における代表的な EEG 所見と一致している。この結果は、クラスタリングにより抽出された主要クラスタが、生理的に妥当な脳波状態を反映していることを示唆する。

一方で、健常者におけるウィンドウの割合が小さいクラスタ 2 およびクラスタ 5 については、それぞれ「デルタ帯域の徐波化による脳機能低下を示唆するパターン」や「低振幅で非同期的な脳活動パターン」と説明されている。これらのクラスタは、健常群では出現頻度が低い一方、認知症群において相対的に多く観測される傾向があり、病的脳波状態を反映している可能性がある。

また、クラスタ 1 およびクラスタ 3 では、高周波数帯域の異常なパワー増大に着目した説明が生成されており、筋電図アーチファクトや筋緊張に起因する信号混入を示唆する内容となっている。このよう

に、LLM は単なる周波数分布の要約にとどまらず、EEG 解析において重要なアーチファクトの存在についても言及しており、専門家の解釈と整合した説明を生成できていることが示された。

5 考察

本研究の定性的分析から、LLM は数値的特徴量に基づく解析結果を、人間に理解可能な自然言語として体系的に説明できることが示された。

従来の機械学習手法では、モデルの出力は数値的指標として提示されることが多く、その判断根拠を人間が理解することは容易ではなかった。これに対し、LLM を用いた説明生成は、数値特徴と既存の神経生理学的知見とを結び付け、解析結果に意味付けを与える役割を果たす。この点は、医療や神経科学など、高い解釈性が要求される分野において特に重要である。

本研究で観察された説明は、EEG の空間的および周波数的特徴に基づいて構成されており、専門家が用いる説明様式と高い親和性を有していた。このことは、LLM がブラックボックス的な予測器としてではなく、解析結果を理解可能な知識へと変換する媒介的な役割を担い得ることを示唆している。

以上より、本研究は、数値的解析と人間の理解との間に存在するギャップを埋める手段として、XAI の有効性を示したものと位置付けられる。特に、LLM を用いた自然言語説明は、教師なし学習により得られた複雑な解析結果を解釈・共有するための一般的な枠組みとして、今後さまざまな応用が期待される。

6 おわりに

本研究では、LLM を用いてクラスタリング結果の説明可能性を向上させる枠組みを提案し、クラスタ重心に基づく神経生理学的に妥当な自然言語説明が生成可能であることを示した。本手法は、数値解析と人間理解の間に存在するギャップを埋めるアプローチとして有効であると考えられる。

今後は、本枠組みを分類手法や他の説明可能 AI と統合するとともに、実データへの適用を通じて、実用性および汎用性の検証を進める予定である。

参考文献

- [1] Priyanshu Pallav and Karunesh Srivastava. Stress analysis of eeg signal using k-means clustering and hybrid classifier. In **2024 2nd International Conference on Advancements and Key Challenges in Green Energy and Computing (AKGEC)**, pp. 1–5, 2024.
- [2] Umut Orhan, Mahmut Hekim, and Mahmut Ozer. Eeg signals classification using the k-means clustering and a multilayer perceptron neural network model. **Expert Systems with Applications**, Vol. 38, No. 10, pp. 13475–13481, 2011.
- [3] 山科勇輔, 須賀圭一, 白井祐典, 市川佳彦. 自然言語での異常解釈: LLM を用いた AI 説明モデルの提案. 言語処理学会 第 31 回年次大会 発表論文集, pp. 2100–2103, 3 2025.
- [4] Chengzong Zhao, Zara Cook, Livia Murray, Jivan Kesan, Nabil Belacel, Sam Doesburg, George Medvedev, Vasily Vakorin, and Pengcheng Xi. Leveraging large language models and fuzzy clustering for eeg report analysis. In **2024 IEEE SENSORS**, pp. 1–4, 2024.
- [5] Andreas Miltiadous, Katerina D. Tzamourta, Theodora Afrantou, Panagiotis Ioannidis, Nikolaos Grigoriadis, Dimitrios G. Tsalikakis, Pantelis Angelidis, Markos G. Tsipouras, Evripidis Glavas, Nikolaos Giannakeas, and Alexandros T. Tzallas. "a dataset of 88 eeg recordings from: Alzheimer's disease, frontotemporal dementia and healthy subjects", 2023.

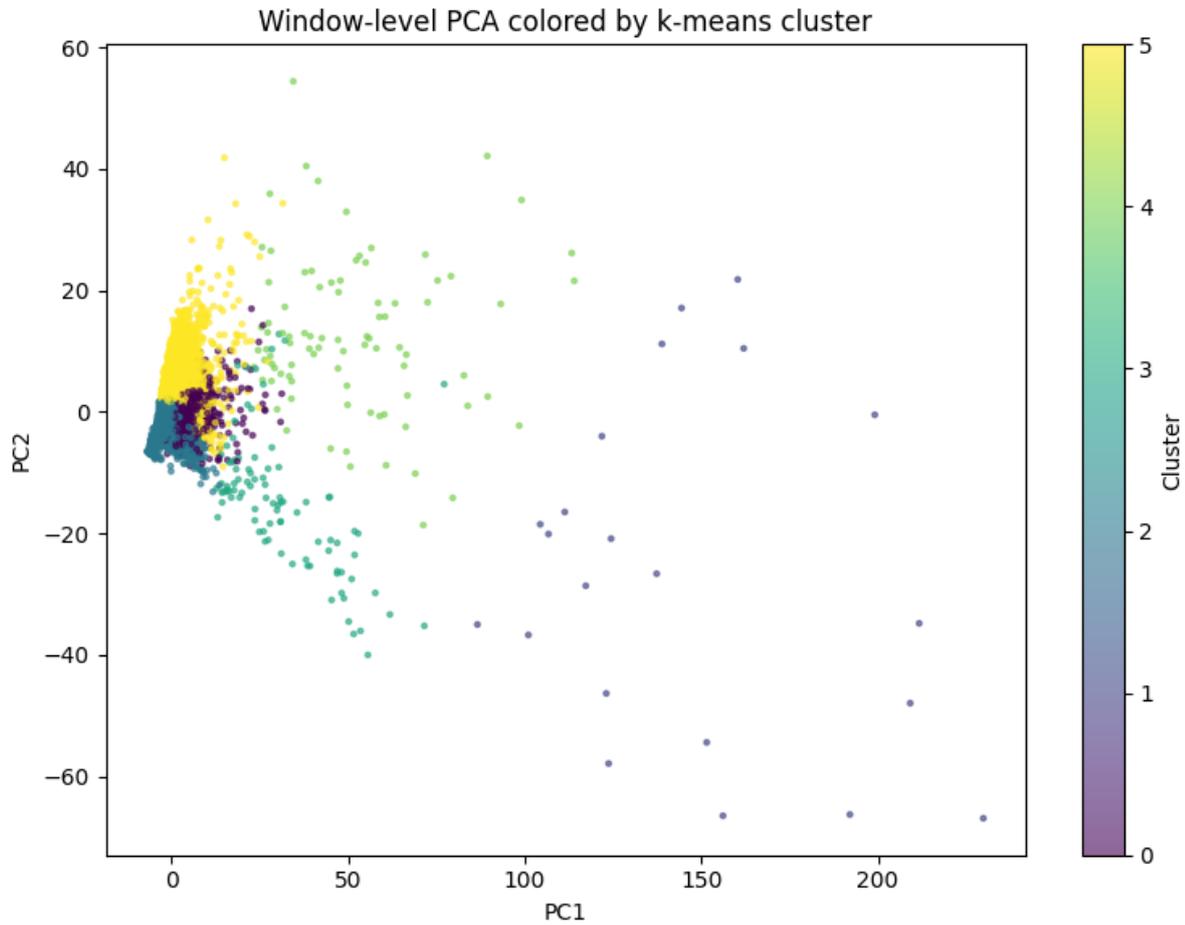


図8 クラスタリング後の PCA 図

表1 クラスタおよび LLM による解説 (要約)

クラスタ ID	解説
0	顕著なアルファ波の亢進を特徴とする, 典型的な閉眼安静覚醒状態の脳波パターン
1	全ての周波数帯域, 特に高周波数帯域においてパワーが極端に増大しており, 筋電図 (EMG) アーチファクトの混入を強く示唆するパターン
2	他の周波数帯域は正常範囲内である一方, デルタ帯域の徐波が選択的に増加している, 脳機能低下を示唆する脳波パターン
3	ベータ波とガンマ波という高周波数成分が極めて顕著に増加し, 軽度の徐波化も伴う, 筋緊張アーチファクトまたは過覚醒状態を示唆するパターン
4	シータ波とデルタ波のパワーが著しく増大しており, 全般的な脳波の徐波化が顕著な, 重度の皮質機能障害を示唆するパターン
5	全ての周波数帯域でパワーが平均を下回る, 低振幅で非同期的な脳活動パターン