

生成系 AI を用いた専門用語の自動抽出と解説

今川 栞成

1. はじめに

近年、ChatGPT や Gemini などの生成系 AI は個人での利用がしやすく、文書要約や画像生成などの機能を日常的に使用する人も多くなっている。井上ら[1]は、e スポーツのストリーマが発した専門用語の解説文を自動生成するシステムを構築している。本稿では、井上らのシステムを参考に、専門分野を学習する際に、障壁となりやすい点の一つに専門用語の多さがある。専門用語の意味をより効率よく把握できるように学習を支援することを目的とし、生成系 AI が自動で専門用語の抽出と解説文の生成を行うシステムを実装し検証を行う。

2. システム構成

本システムは、生成系 AI (Google Gemini) を使い専門用語の抽出と解説文の生成を自動で行う。本システムは、音声を入力として読み込み文字起こしを行い、単語群を生成する。次に、単語群より1語、連続する2〜3単語をキーワードとし、各キーワードと単語群との関連度 (コサイン類似度) を計算する。関連度の高い上位15個のキーワードを専門用語候補とする。最後に、生成系 AI が専門用語であると判断したものを専門用語とする。併せて、解説文を生成する。図1に入力となる音声から、出力となる専門用語とその解説文を生成するまでの過程を示す。

ステップ1：音声からの文字起こし
「医学では人体の構造と機能、そして疾病の診断と治療を研究します。診断の分野では CT や MRI といった画像診断技術が用いられ、臓器や組織の状態を詳細に…」

ステップ2：キーワード (1単語、連続する2〜3単語)
人体、診断、研究、CTやMRI、画像診断技術…

ステップ3：専門用語候補 (関連度の高いキーワード)
CTやMRI、薬物療法、疾病、構造、放射線…

ステップ4：専門用語
プロンプト「以下の単語リストから、一般的な単語ではなく専門用語のみを抽出してください。」
CTやMRI、薬物療法、疾病、再生、構造、放射線、個別化、マーカー、有無

ステップ4：解説
プロンプト「〇〇について、50字程度で簡単に説明してください」
(1) CTやMRI：
CT：X線での体の断面を撮影。骨や出血の診断に。
MRI：磁場と電波で体の内部を詳細に撮影。…
(2) 薬物療法：
薬物療法とは、病気の治療や症状の緩和に薬を用いる方法です。原因を取り除いたり、体の機能…

図 1 解説文生成までの過程と例

3. 実験

20 個の異なる分野 (情報学や医学、法学など) に対して、30 秒〜1 分程度の説明音声を用意し、専門用語の抽出と各専門用語において 50 字程度の解説文を生成する。抽出された各専門用語に対して「専門用語の抽出は正確であるか」、「生成された専門用語の解説は正確であるか」を 10 人の被験者にそれぞれ 1〜5 の 5 段階で回答してもらい、最後に「このようなシステムは専門分野学習時に有用かどうか」を 1〜10 の 10 段階で評価してもらう。

4. 結果

図 2 と図 3 に結果を示す。専門用語と解説の正確さは分野を通して顕著な差はみられず、平均はそれぞれ 4.18 と 4.44 であった。また、本システムを専門分野の学習時の支援としての有用性については平均は 8.9 であり、本実験の被験者に対しては好評であった。

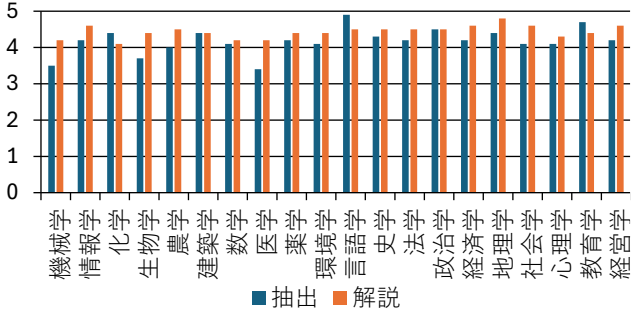
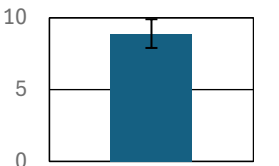


図 2 専門用語と解説の正確さ

5. おわりに

専門用語と解説の正確さ、有用性がともに 高評価であることから実時間で処理ができれば、専門分野の学習に有用な方法となる可能性がある。一方で、本実験の被験者は学生のみであったため、より一般的な評価を得るために、より幅広い分野や年齢層を被験者とした検証が必要である。

図 3 有用性



[参考文献]

[1] 井上奏太, 梅原旅詩, 川合康央, ChatGPT を用いた e スポーツ専門用語自動解説システムの開発と評価, 情報処理学会 2025, 1P-74, 2025