

【背景・目的】2022年以降、AIは急速に進化し、マルチモーダルな利用が可能となった。しかし、AIへの指示は依然として言語が中心であり、その精度はユーザ自身の言語化能力に依存している。本研究では、図1のようにナレッジグラフとAIを活用して関心分野の拡張・収束を行い、「思考の拡散」をエントロピーを用いて定量的に評価することを目指した。

【検証方法】検証の題材として、「レアアース」に関する5つの異なる観点(1.物質科学、2.貿易、3.環境問題、4.地質学、5.欧州の情勢)の文書を用いた(図2、表1に対応)。具体的な手順は以下の通りである。まず、AIに文献を入力して関係性データを出力させ、そこから選択した単語を起点にグラフの拡張を2回行った。次に、拡張された全関係性データを再びAIに入力して要約・文章化し、その埋め込みベクトル(3072次元)を取得した。評価は二つの観点で行った。第一に、特異値分解(SVD)、ランダム射影、主成分分析(PCA)、多次元尺度構成法(MDS)の4手法を用いて埋め込みベクトルを視認可能な3次元まで圧縮し、各観点の類似性を評価した。第二に、AIが出力した関係性データの「グラフ構造的な複雑性」と、グラフ拡張による「話題の拡張性」をエントロピーによって導出した。前者には von Neumann entropy [1]を用い、後者はノードの単語を BERTopic [2]によりトピック分類し、各トピックの entropy の合計値を指標として検証した。

【結果】まず、次元圧縮による埋め込みベクトルの可視化結果を示す。図2は、PCAにより文章ベクトルを3次元まで圧縮・可視化したものである。1(物質科学)と4(地質学)、2(貿易)と3(環境問題)がそれぞれ重なってプロットされており、内容の類似性が示唆された。また、5(欧州の状況)の話題は他4つのプロットと離れており、埋め込みベクトルの次元圧縮による可視化が、内容の共通性や関連性の確認

に有益であることが分かった。次に、グラフ拡張前後のエントロピーの変化を表1に示す。グラフ構造のエントロピーと、話題の拡張性を示したエントロピーの双方において、グラフ拡張後に値が増加していることが確認できた。これにより、AIを利用した思考の拡張を、エントロピーを用いた定量評価によって情報量の増加として示すことが可能となった。

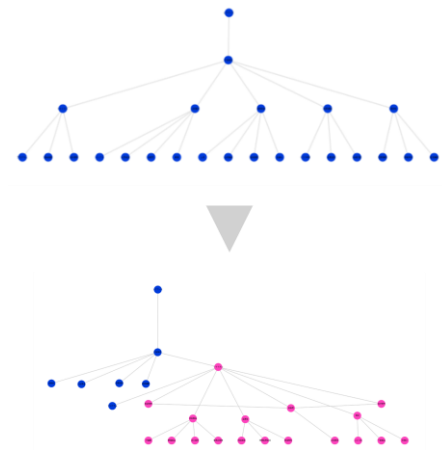


図1. グラフの拡張と削除

表1. 拡張前後の entropy

|   | Graph Entropy |        | Totall Topic Entropy |         |
|---|---------------|--------|----------------------|---------|
|   | 前             | 後      | 前                    | 後       |
| 1 | 4.4002        | 5.1224 | 12.1755              | 22.3912 |
| 2 | 3.8259        | 5.0294 | 8.0677               | 19.9057 |
| 3 | 5.6177        | 6.1079 | 29.3982              | 43.0739 |
| 4 | 5.0524        | 5.7305 | 21.1112              | 28.6465 |
| 5 | 4.5499        | 5.3100 | 15.1276              | 24.6525 |

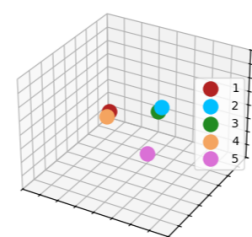


図2. PCAの次元削減による可視化

[1] Passerini, Filippo, and Simone Severini. "The von Neumann entropy of networks." arXiv preprint arXiv:0812.2597 (2008).  
[2] Grootendorst, Maarten. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." arXiv preprint arXiv:2203.05794 (2022).