人工知能学会 インタラクティブ
情報アクセスと可視化マイニング研究会(第35回)
SIG-AM-35-01

LLMを用いた推薦リスト生成のための ペアワイズ比較・集約手法についての予備的検討

Preliminary Study on Pairwise Comparison and Aggregation Methods for Recommendation List Generation Using LLMs

谷 知拓[1*]　柴田 祐樹[2]　高間 康史[2]
Tomohiro Tani[1]　Hiroki Shibata[2]　Yasufumi Takama[2]

[1] 東京都立大学 システムデザイン学部
[1] Faculty of Systems Design, Tokyo Metropolitan University
[2] 東京都立大学大学院 システムデザイン研究科
[2] Graduate School of Systems Design, Tokyo Metropolitan University

Abstract: This paper investigates pairwise comparison and aggregation approaches for generating recommendation lists using Large Language Models (LLMs) in a zero-shot and scalable manner. When dealing with a large number of candidate items that cannot be evaluated by LLMs in a single prompt, it becomes necessary to divide the task into multiple prompts and aggregate the results. The proposed method prompts LLMs to estimate preference order between two items, and generates recommendation lists through Bradley-Terry-Luce (BTL) aggregation. This paper reports the results of preliminary experiments and discusses the impact of the number of item pairs and LLM input batch size on the consistency and accuracy of recommendations.

1 はじめに

本稿では，LLM を用いたゼロショット・スケーラブルな推薦リスト生成手段として，ペアワイズ比較を集約するアプローチに着目し，予備実験を行った結果について報告する．

近年，大規模言語モデル（Large Language Model: LLM）の急速な発展により，自然言語処理の様々なタスクにおいて革新的な成果が得られている．推薦システムの分野においても，LLM を活用した新たなアプローチが注目を集めており，従来の協調フィルタリングや内容ベースフィルタリングとは異なる観点から，ユーザの嗜好を理解し推薦を生成する手法が提案されている [2, 3, 4, 5]．特に，LLM の持つゼロショット学習能力を活用することで，事前の学習データや特徴量エンジニアリングを必要とせずに，テキスト記述のみから推薦を生成できる可能性が示されている [1]．

しかしながら，LLM を用いた推薦システムの実用化においては，いくつかの技術的課題が存在する．特に，推薦候補となるアイテム数が多い場合，すべてのアイテムを一度に LLM に入力することは，入力トークン数の制限により困難である．この問題に対処するため，候補アイテムを複数のバッチに分割し，段階的に処理を行うアプローチが考えられるが，その際に候補アイテムのバッチへの分割，LLM の出力の統合をどのように行い，最終的な推薦リストを生成するかが重要な課題となる．

本稿では，この課題に対するアプローチとして，ペアワイズ比較に基づく手法に着目する．具体的には，アイテムのペアごとに LLM に選好関係を評価させ，得られた比較結果を Bradley-Terry-Luce（BTL）モデル [8] を用いて集約することで，全体的な推薦リストを生成する手法を提案する．ペアワイズ比較アプローチは，各比較において考慮すべき情報量を限定できるため，LLM の入力制限に対して頑健であり，また比較の並列処理が可能であるという利点を持つ．

本稿では，提案手法の予備的な検討として，ペア数および LLM への入力バッチサイズが推薦の一貫性と精度に与える影響について評価実験を行った結果を報告する．実験では，実際の商品レビューデータを用いて，異なる設定での推薦リスト生成を行い，その性能を比較分析する．実験の結果，ペア数が増加するにつれて推薦精度と一貫性が向上する傾向が確認されたが，50 ペアを超えると性能が飽和することが観測された．

*連絡先：東京都立大学システムデザイン学部
〒 191-0065 日野市旭が丘 6-6
E-mail: tani-tomohiro@ed.tmu.ac.jp

また，バッチサイズについては，バッチサイズ 2 の純粋なペアワイズ比較において最も高い一貫性が得られたことを報告する．

2 関連研究

2.1 LLM を用いた推薦システム

大規模言語モデルを推薦システムに応用する研究は，近年活発に進められている．特に，事前の学習データなしに推薦を行うことが可能なゼロショット学習による推薦生成 [1] は，コールドスタート問題の解決やビックデータを必要としない柔軟な推薦システムの構築手段として，期待が寄せられている．また，協調フィルタリングの概念を LLM に組み込んだ手法 [2, 6] や，グラフ構造を活用した手法 [3]，生成的アプローチによる個人化推薦 [4, 5] など，様々な観点からの研究が進められている．

しかし，既存の LLM ベース推薦手法の多くは，候補アイテム数が限定的な場合を想定しており，大規模なアイテム集合に対するスケーラビリティの課題が残されている [7].

2.2 ペアワイズ比較に基づく選好学習

2 つの選択肢を比較するペアワイズ比較は，多数の選択肢を同時に評価するよりも認知負荷が低く，より正確な判断が可能であることが知られている [9]．LLM においても，ペアワイズ比較は有効なアプローチとして注目されている．複数のアイテムを一度に評価させる場合と比較して，2 つのアイテムの相対的な優劣を判断させることで，より一貫性のある選好情報を獲得できることが報告されている [1, 7].

Bradley-Terry-Luce（BTL）モデルは，ペアワイズ比較結果から全体的なランキングを推定する統計モデルである [8]．各アイテムに潜在的な強度パラメータを仮定し，アイテム i がアイテム j より選好される確率を以下のように表現する：

$$P(i \succ j) = \frac{\pi_i}{\pi_i + \pi_j} \tag{1}$$

ここで，π_i はアイテム i の強度パラメータであり，観測された比較結果から最尤推定や反復アルゴリズムにより求められる．例えば，3 つのアイテム A，B，C に対して「A が B に勝つ」「B が C に勝つ」「A が C に勝つ」という比較結果が得られた場合，これらの観測確率を最大化するように各アイテムの強度パラメータ π_A, π_B, π_C を推定し，その値に基づいて全体のランキング（この例では A > B > C）を決定する．

BTL モデルは，トーナメントやスポーツランキング等の分野で広く利用されている．

3 提案手法

3.1 システム概要

本稿では，LLM を用いたスケーラブルな推薦システムを実現するため，ペアワイズ比較と BTL モデルによる集約に基づく推薦システムを提案する．提案システムは，以下の 3 つの主要コンポーネントから構成される：(1) 候補アイテムのペアワイズ比較を行う LLM モジュール，(2) 比較結果を集約する BTL モデル，(3) 最終的な推薦リストを生成するランキングモジュール．

システムの処理フローは以下の通りである．まず，ユーザの嗜好情報と候補アイテム集合を入力として受け取る．次に，候補アイテム集合からアイテムペアを生成し，各ペアについて LLM に選好判定を要求する．LLM は，ユーザの嗜好情報に基づいて，どちらのアイテムがより適切かを判定する．得られた比較結果は BTL モデルに入力され，各アイテムの強度パラメータが推定される．最後に，推定された強度パラメータに基づいてアイテムをソートし，推薦リストを生成する．

本システムの特徴は，アイテム数が多い場合でも，ペアワイズ比較により処理を分割できる点にある．すべてのアイテムを一度に LLM に入力する方法と比較して，本システムでは各比較を独立に実行できるため，並列処理が可能であり，スケーラビリティが向上する．

3.2 ペアワイズ比較の実施

ペアワイズ比較では，候補アイテム集合から 2 つのアイテムを選択し，LLM に対してどちらがユーザにとって適切かを判定させる．具体的には，以下の形式で LLM にプロンプトを与える：

> ユーザの嗜好情報：[ユーザの過去の購買履歴やレビュー]
>
> 以下の 2 つのアイテムのうち，このユーザにより適していると思われるものを選択してください．
>
> アイテム A：[アイテム A の特徴・説明]
>
> アイテム B：[アイテム B の特徴・説明]

ペアの生成方法については，候補アイテム集合から n 個のペアを，重複して同じペアが選ばれないようにサンプリングする．ペア数 n は重要なパラメータであり，多すぎると LLM へのクエリ数が増加しコストが

上昇する一方，少なすぎると推薦精度が低下する可能性がある．

バッチサイズは，一度の LLM クエリで比較するアイテム数を制御する．$b = 2$ の場合は純粋なペアワイズ比較となり，$b > 2$ の場合は複数ペアを統合したアイテム群での比較となる．異なるバッチ間では LLM の判断に矛盾が生じる可能性がある．BTL モデルはペア間で選好順序に矛盾がある場合も処理可能であるが，バッチサイズは推薦の一貫性に影響する重要なパラメータと考える．

3.3 BTL 集約による推薦リスト生成

LLM によるペアワイズ比較結果を集約するため，BTL モデルを使用する．本稿では計算効率と数値安定性の観点から，対数スコア $s_i = \log \pi_i$ を用いた定式化を採用する．この関係を式 (1) に代入すると，次式が得られる：

$$P(i \succ j) = \frac{1}{1 + \exp(s_j - s_i)} \tag{2}$$

ロジスティック関数を用いた表現により，スコア差が大きいほど選好確率が 1 に近づき，また勾配に基づく最適化が容易になる．

3.3.1 パラメータ推定アルゴリズム

本研究の実装では，確率的勾配降下法（SGD）を用いてスコアパラメータを推定する．観測されたペアワイズ比較結果の集合を $\mathcal{D} = \{(w_k, l_k)\}$ とする．ここで，w_k は k 番目の比較における勝者，l_k は敗者を表す．

各比較 (w_k, l_k) に対して，以下の更新式でスコアを調整する：

$$p_k = P(w_k \succ l_k) = \frac{1}{1 + \exp(s_{l_k} - s_{w_k})} \tag{3}$$

$$s_{w_k} \leftarrow s_{w_k} + \alpha(1 - p_k) \tag{4}$$

$$s_{l_k} \leftarrow s_{l_k} - \alpha(1 - p_k) \tag{5}$$

ここで，α は学習率（デフォルト値：0.01）である．この更新により，勝者のスコアは増加し，敗者のスコアは減少する．更新量 $(1 - p_k)$ は，現在のモデルによる予測確率と実際の結果（勝者の勝利確率=1）との差に相当する．

3.3.2 不確実性の定量化

BTL モデルでは，アイテムペア (i, j) の比較における不確実性を以下のように定量化できる：

$$U(i, j) = 2 \times \min\{P(i \succ j), P(j \succ i)\} \tag{6}$$

この不確実性指標は，両アイテムのスコアが近い場合に最大値 1 をとり，スコア差が大きい場合に 0 に近づく．本システムでは，この不確実性を利用してアクティブサンプリングを行い，不確実性の高いペアを優先的に LLM に評価させることで，クエリ効率を向上させている．

3.3.3 収束性と計算量

提案手法では，全比較データに対して 100 回の反復を行うことで収束を図る．1 回あたりの計算量は $O(|\mathcal{D}|)$ であり，全体の計算量は $O(100 \times |\mathcal{D}|)$ となる．実験では，$|\mathcal{D}|$ は最大 100 ペアであり，計算は数ミリ秒で完了した．

最終的に，推定されたスコア $\{s_i\}$ に基づいてアイテムを降順にソートすることで，推薦リストを生成する．この手法により，限られた数のペアワイズ比較から，全アイテムの相対的な順位を効率的に推定することが可能となる．

4 評価実験

4.1 データセット

Amazon Review Dataset 2023[1][7] の Movies and TV カテゴリを使用して評価実験を行った．このデータセットは，Amazon プラットフォーム上で公開されている映画・TV 番組に関するユーザレビューを含んでおり，推薦システムの評価に広く利用されている．Movies and TV カテゴリを選択した理由は，LLM が事前学習において映画や TV 番組に関する豊富な知識を獲得していることが予想され，アイテムの内容を理解した上での推薦が期待できるためである．

データセットから，各ユーザに対して正解（実際に高評価を付けたアイテム）1 件と，ランダムに選択した 19 件の計 20 件を候補アイテムとして使用した．この設定により，推薦タスクは 20 個のアイテムから最適なものを識別する問題となる．候補アイテム数を 20 件に限定した理由は，ペアワイズ比較の効果を明確に評価するためと，実験の計算コストを現実的な範囲に抑えるためである．

各アイテムについては，タイトル，ジャンル（カテゴリ），平均評価およびレビュー件数，特徴，製品説明等のメタデータを使用し，LLM がアイテムの内容を理解できるようにした．ユーザの嗜好情報としては，最

[1] https://amazon-reviews-2023.github.io/

近評価したアイテム（2-5 件程度）のメタデータを要約したテキストを使用し，各ユーザの好みを LLM に伝えられるようにした．

4.2 実験設定

4.2.1 評価指標

推薦システムの性能を多角的に評価するため，以下の指標を使用する．

(1) 推薦精度指標

- **Hit@10**：正解アイテムが推薦リストの上位 10 位以内に含まれる割合．この指標は推薦システムが関連アイテムを上位に配置できているかを評価する．
- **平均逆順位（Mean Reciprocal Rank, MRR）**：正解アイテムの順位の逆数の平均値．より上位に正解が現れるほど高い値となる．
- **平均順位**：正解アイテムが推薦リスト中に現れる順位の平均値．20 個の候補中での絶対的な位置を示す．

(2) 推薦の一貫性指標

同一ユーザ・同一条件で複数回（5 回）推薦を行った際の結果の安定性を以下の指標で評価する：

- **Jaccard 類似度**：2 つの推薦リストの上位 10 アイテム集合に対する類似度．$J(A,B) = |A \cap B|/|A \cup B|$ で計算され，0 から 1 の値をとる．
- **Spearman 順位相関係数**：2 つの推薦リストの上位 10 アイテムのうち，共通するアイテムの順位相関．順位の一致度を-1 から 1 の範囲で評価する．

4.2.2 バッチサイズに関する実験

バッチサイズ b を $\{2, 5, 10\}$ と変化させて得られたペアの集合から，ランダムに 50 ペアを抽出し，推薦の一貫性に与える影響を評価した．バッチサイズは LLM への一度の入力で比較するアイテム数を制御するパラメータであり，b 個のアイテムから $\binom{b}{2}$ 個のペアワイズ比較が生成される．

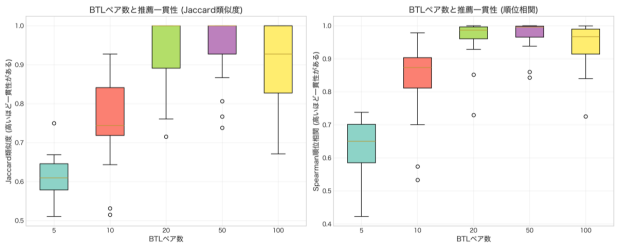


図 1: ペア数と推薦の一貫性の関係．(a) Jaccard 類似度，(b) Spearman 順位相関係数．

4.2.3 実験手順

各実験は以下の手順で実施した：

1. 推薦対象（U）の各ユーザについて正解 1 件＋ランダム 19 件の候補アイテムを準備．
2. アクティブサンプリングにより，不確実性（式 (6)）の高いアイテム集合（サイズ=b）を選択．
3. 選択されたアイテム集合からアイテムペアを生成して LLM に選好順序を判定させ，3.3 節で述べた手順により，BTL モデル（反復 100 回）により各アイテムの s_i を推定．
4. s_i に基づいてアイテムをソートし，推薦リストを生成．
5. 評価指標を計算．

すべての実験条件で同一のユーザ・アイテム集合を使用することで，公平な比較を実現した．

4.3 実験結果

4.3.1 ペア数が精度・一貫性に与える影響

BTL モデルに入力するペア数が推薦性能に与える影響を評価するため，ペア数を $\{5, 10, 20, 50, 100\}$ の 5 段階で変化させて実験を行った．バッチサイズは 2 に固定し，50 ユーザに対して各条件で 5 回の試行を実施した．

表 1 に，ペア数と推薦精度の関係を示す．Hit@10 は，ペア数の増加に伴って向上する傾向が観察されたが，ペア数 100 ではわずかに低下する傾向が見られた．同様に，平均逆順位（MRR）と平均順位においても，ペア数 50 付近で最良の性能を示し，それ以上では性能が飽和または低下する傾向が確認された．

図 1 に，同一条件での 5 回の試行結果に対する Jaccard 類似度と Spearman 順位相関係数を示す．両指標ともに，ペア数の増加に伴って一貫性が向上する傾向が観察された．特に Spearman 順位相関係数では，より滑

表 1: ペア数と推薦精度指標の統計サマリ（50 ユーザの平均値±標準偏差）

ペア数	Hit@10	MRR	平均順位
5	0.42 ± 0.18	0.18 ± 0.11	8.2 ± 3.4
10	0.58 ± 0.16	0.27 ± 0.13	6.5 ± 2.8
20	0.65 ± 0.14	0.33 ± 0.12	5.8 ± 2.5
50	0.72 ± 0.12	0.38 ± 0.11	5.2 ± 2.2
100	0.70 ± 0.13	0.36 ± 0.12	5.4 ± 2.3

表 2: ペア数と推薦の一貫性指標の統計サマリ（50 ユーザの平均値±標準偏差）

ペア数	Jaccard	Spearman
5	0.35 ± 0.12	0.42 ± 0.15
10	0.48 ± 0.10	0.56 ± 0.12
20	0.58 ± 0.09	0.65 ± 0.10
50	0.68 ± 0.08	0.73 ± 0.09
100	0.66 ± 0.09	0.71 ± 0.10

らかな上昇曲線が得られた．これは，順位情報を考慮した指標であるため，推薦リストの順位の安定性をより適切に評価できているためと考える．

ペア数が 50 を超えると，一貫性の低下が観察された．この現象は，過度に多くのペアワイズ比較を行うことで，ノイズや矛盾が蓄積される「過学習」に類似した現象が生じている可能性を示唆しており，推薦精度の低下の一因になったと考える．

4.3.2 バッチサイズが精度・一貫性に与える影響

次に，BTL に入力するペア数を 50 に固定した状態で，バッチサイズを変化させた際の影響を分析した．バッチサイズは，一度のプロンプトで比較するアイテム数を制御するパラメータであり，計算効率と推薦品質のトレードオフを決定する重要な要素である．

図 2 に，バッチサイズと推薦の一貫性指標の関係を示す．バッチサイズが 2 の場合に最も高い一貫性を示し，バッチサイズの増加に伴って一貫性が低下する傾向が観察された．

この結果は，小さなバッチサイズでは LLM がより明確な選好判断を行えることを示唆している．2 つのアイテムの直接比較は最もシンプルなタスクであり，LLM は一貫した判断を下しやすい．一方，多数のアイテムを同時に順位付けする必要がある場合，タスクの複雑性が増し，出力の変動が大きくなると考える．

また，正解アイテムの順位の標準偏差およびレンジ（最大値と最小値の差）（図 2 下段）も，バッチサイズの増加とともに単調に増加する傾向を示し，推薦結果の不安定性が顕著となった．

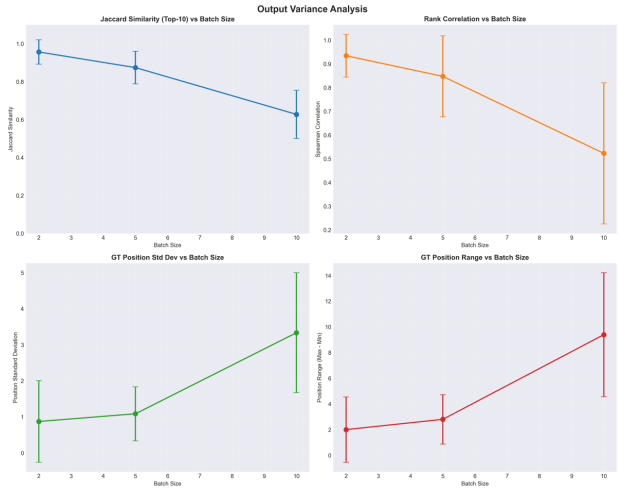


図 2: バッチサイズと推薦の一貫性の関係．上段左：Jaccard 類似度，上段右：Spearman 順位相関，下段左：正解順位の標準偏差，下段右：正解順位のレンジ（最大値-最小値）．エラーバーはユーザ間のばらつき（標準偏差）を示す．

4.4 考察

4.4.1 実験結果の解釈

本実験の結果から，LLM を用いたペアワイズ比較に基づく推薦システムにおいて，ペア数とバッチサイズが推薦品質に与える影響について，以下の重要な知見が得られた．

第一に，ペア数と推薦精度の関係において，単純な比例関係ではなく，ある閾値（本実験では 50 ペア）で性能が飽和する傾向が確認された．これは，BTL モデルが一定数以上の比較データから十分な統計的情報を抽出できることを示している．一方で，ペア数が 100 を超えると精度が若干低下する現象は，推移律違反の増加と関連していると考える．推移律違反などの矛盾した判断については今後詳細に分析する予定であるが，LLM の出力に内在する確率的な揺らぎが，大規模な比較集合において矛盾として顕在化し，BTL 集約の品質を低下させている可能性がある．

第二に，バッチサイズと推薦の一貫性の間には負の相関が観察された．バッチサイズ 2（純粋なペアワイズ比較）において最も高い一貫性が得られた理由として，タスクの認知的複雑性が考えられる．人間の意思決定研究においても，選択肢が増加すると判断の一貫性が低下することが知られており，LLM も同様の傾向を示すことが示唆される．また，大きなバッチサイズでは，アイテム間の相対的な特徴の差異が希薄化し，順位付けの基準が不安定になる可能性がある．

5 おわりに

本稿では，LLM を用いたゼロショット・スケーラブルな推薦リスト生成のために，ペアワイズ比較結果を BTL モデルで集約する手法を検討した．

評価実験では，精度指標（Hit@10，MRR，平均順位）と一貫性指標（Jaccard 類似度，Spearman 順位相関）がペア数，バッチサイズによりどのように変化するかを調査した．その結果，ペア数が増加するにつれて推薦精度と一貫性が向上する傾向が確認されたが，50 ペアを超えると性能が飽和することが観測された．また，バッチサイズ 2 において最も高い一貫性が得られた．

今後は，推移率違反などの矛盾の発生について調査する他，既存の協調フィルタリングやコンテンツベースフィルタリング手法との精度や計算効率の比較検証を進め，LLM を用いたゼロショット推薦手法の有効性を調査する予定である．

謝辞

本研究の一部は，JSPS 科研費 22K19836，23K24953 の助成を受けたものです．

参考文献

[1] Hou, Y., Zhang, J., Lin, Z., Lu, H., Xie, R., McAuley, J., Zhao, W.X.: Large Language Models for Zero-Shot Recommender Systems, *ECIR2024*, pp. 364-381 (2023)

[2] Yao, S., Wu, L., Guo, Q., Hong, L., Li, J.: Collaborative Large Language Model for Recommender Systems, *WWW'24*, pp. 3162–3172 (2024)

[3] Wei, W., Ren, X., Tang, J., Wang, Q., Su, L., Cheng, S., Wang, J., Yin, D., Huang, C.: LLMRec: Large Language Models with Graph Augmentation for Recommendation, *WSDM'24*, pp. 806–815 (2024)

[4] Li, J., Zhang, W., Wang, T., Xiong, G., Lu, A., Medioni. G.: GPT4Rec: A Generative Framework for Personalized Recommendation and User Interests Interpretation, *SIGIR eCom 2023* (2023)

[5] Ngo, H., Nguyen, D.Q.: RecGPT: Generative Pre-training for Text-based Recommendation, *ACL2024*, pp. 302–313 (2024)

[6] Kim, S., Kang, H., Choi, S., Kim, D., Yang, M., Park, C.: Large Language Models meet Collaborative Filtering: An Efficient All-round LLM-based Recommender System, *KDD'24*, pp. 1395–1406 (2024)

[7] Hou, Y., Li, J., He, Z., Yan, A., Chen, X., McAuley, J.: Bridging Language and Items for Retrieval and Recommendation, *arXiv preprint*, arXiv:2403.03952 (2024)

[8] Bradley, R.A., Terry, M.E.: Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons, *Biometrika*, Vol. 39, No. 3/4, pp. 324–345 (1952)

[9] Guo, S., Sanner, S.: Real-time Multiattribute Bayesian Preference Elicitation with Pairwise Comparison Queries, *AISTATS'2010*, pp. 289-296 (2010)

自動車灯火装置における故障推論のための構造と故障の概念精緻化の検討

A Study on Knowledge Models for Fault Inference and Refining the Concept of Failure in Automotive Lighting Devices

島村 佳周 [1] *　神田 脩太朗 [1]　笹嶋 宗彦 [1]
Yoshinari Shimamura[1]　Syutaro Kanda[1]　Munehiko Sasajima[1]

[1] 兵庫県立大学
[1] University of Hyogo

Abstract: In automotive design, it is difficult to comprehensively infer possible failures. In this study, the authors have been studying fault inference aimed at supporting that task. In this paper, we modified the knowledge model structure and verified its operation to address the issue that the fact that content which should have been defined as the structure was included in parts names was hindering fault inference in the knowledge model of previous research. Furthermore, to address the issue of being unable to handle failure causes such as aging deterioration, we refined the concept of failure.

1　はじめに

本研究では，オントロジー工学の技術を用いることにより，機能を発揮する人工物において起こり得る故障をなるべく多く推論するシステムを構築することを目的とする．

自動車や時計などの機能を発揮する人工物は，使用するにつれて故障が発生する．故障を診断したり，修理したり，防止したりするためには，人工物の設計者は，対象となる部品が故障した際における他の部品への影響や，その人工物全体への影響を推論する必要がある．例えば，自動車の設計において，設計者は設計変更業務を行うために，設計変更を行う部品の過去の故障事例に関する情報を収集する．そして，その情報から設計変更を行った際の，他の部品や自動車全体への影響を網羅的に推論して列挙している [1].

先行研究 [1] では，自動車のパワートレイン部を題材に，オントロジーと機能分解木を用いて構築した知識モデルによって部品と機能の観点から，部品の故障知識を網羅的に導き出す故障推論の有効性を確認した．しかし，自動車における基本メカニズム別の分類では，パワートレイン系，ブレーキ系，吸排気系など様々な系統が存在する [2]．そのため，先行研究の知識モデルをこれらの系に拡張した場合に，故障推論が正しく動作するか検討する必要がある．

*連絡先：島村佳周，兵庫県立大学大学院
〒 651-2197 兵庫県神戸市西区学園西町 8 丁目 2-1，078-794-5794
E-mail:syo2001rimuru.7@outlook.jp

また，先行研究 [3] では，故障推論に必要な概念定義の仕方に関する分析を行い,「不具合」概念の精緻化を行った．そして，先行研究 [1] に加えて，経年劣化のような長期間で蓄積される故障概念を表現可能とする「不具合」概念の定義の仕方を提案した．しかし，実際に故障推論を実行した結果，表面的な不具合のみが推論され，経年劣化のような深層的な不具合が推論されていない場合があった．そのため，不具合に関する定義を再検討し，故障推論が適切に動作するかを改めて検証する必要がある．

以上の背景に基づき，本研究では，自動車の灯火装置に関するオントロジーと，先行研究 [1] で構築した自動車のパワートレイン部に関するオントロジーを統合し，知識モデルの拡張を行う．そのうえで，拡張過程で明らかとなった問題点について考察する．さらに，故障（不具合）の分類をより精緻化し，故障がどのような過程で発生するのかを定義した「自動車故障過程オントロジー」を構築することで，経年劣化のような故障原因に関するオントロジーについても考察を行う．

2　故障推論における知識モデルの構築

筆者らは，知識モデル構築の対象として自動車の灯火装置を選択し，故障推論におけるオントロジー，機能分解木を構築した．灯火装置には，ヘッドランプやテールランプなどがあり，主に目の届く箇所に取り付けられているため，モデル化の対象としやすい．また，

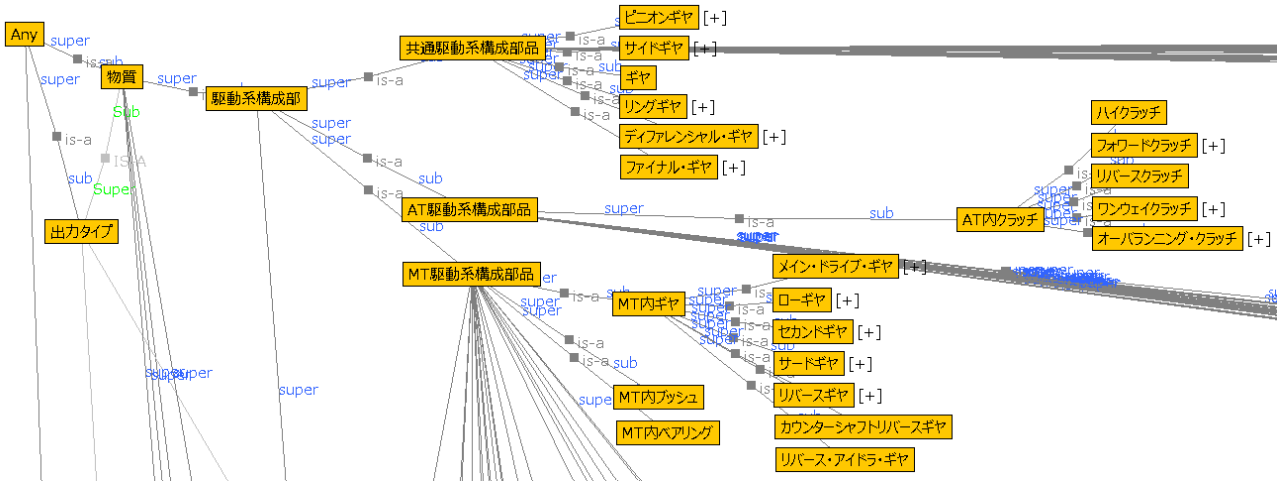


図 1: パワートレイン部オントロジーにおける「内」を含む単語の定義

各ランプの機能は主に発光することであるため，単純に記述できると予測した．一方，先行研究 [1] の対象である自動車のパワートレイン部は，普段目の届きづらい部分にあり，様々な機能があることからシステムが複雑になり，部品の数も非常に多い．以上の考察をふまえて，本研究では，パワートレイン部と灯火装置を対象として，知識モデルの拡張を試みる．また，灯火装置を対象として，故障概念についての精緻化を試みる．

本研究で用いたオントロジーは，対象とする世界を説明するために必要な「概念」を定義する辞書のようなものであり,「概念クラス」とそれらの関係を表す「意味リンク」で構成されている [4]．本研究では，オントロジーを構築する際に，法造 [5] を用いた．また，オントロジー構築の際には，自動車に関する専門書 [6] から知識を抽出して反映させた．その結果，現在の灯火装置オントロジーの基本概念数は，90 個となっている．

また，本研究で用いた機能分解木は，実現したい機能について，それを達成できる部分機能の系列に展開した知識モデルのことであり [7]，本研究では FWTEditor というソフトウェアを用いて記述する．機能分解木の構築においては，オントロジーの構築時と同様に，自動車に関する専門書から知識を抽出し，反映する．

3 故障推論

故障推論とは，ある正常な機能を置き換えた場合に，他のどの機能に影響が及ぶかを網羅的に導出することである [1]．また，機能分解木とオントロジーでベテラン設計者が持つ知識をモデル化し，部品の機能から起こり得る故障知識を導出できるようにするシステムである．飯田らは先行研究 [1] で，古崎らによるオントロジーを用いて分野横断的にキーワードを検索する方法論 [8] を参考に，故障推論を行う方法について設計と実装を行っている．その故障推論の方法については，以下の通りである [1]．

1. 作成したオントロジーを RDF ファイルとしてエクスポートし，サーバ内のデータベースに格納する．
2. FWTEditor 上で機能分解木のノードを 1 つ選択すると，選択したノードに記述されている文章が単語分割され，単語が抽出される．
3. その単語を用いて法造上のオントロジーへ OR 検索が行われ，単語からどのような故障が起こり得るかをシステム的に導出する．OR 検索とは，いずれかの単語を含む定義内容をオントロジー上で検索することである．検索の際には，設計変更の対象となる部品について，その部品を構成する副部品や，その部品の機能が伝搬する他の部品を探索する．
4. 故障名称にたどり着いたら推論を終了し，円状のグラフ，または表形式で結果を表示する．

4 故障推論における知識モデルの拡張

本研究では，故障推論における知識モデルの拡張の対象として，自動車の灯火装置オントロジーおよび，先行研究 [1] で構築した自動車のパワートレイン部オントロジーを用いる．これら 2 つのオントロジーをサーバ内のデータベースに登録し，故障推論を実行した．本章では，故障推論の実行時に生じた問題点を示し，それに対して実施した解決方法および考察について述べる．

4.1 知識モデル拡張時における問題点

灯火装置に関する機能分解木において，特定の機能ノードを選択し，故障推論を実行した．その際，本来は灯火装置内で発生する故障のみが出力されることが

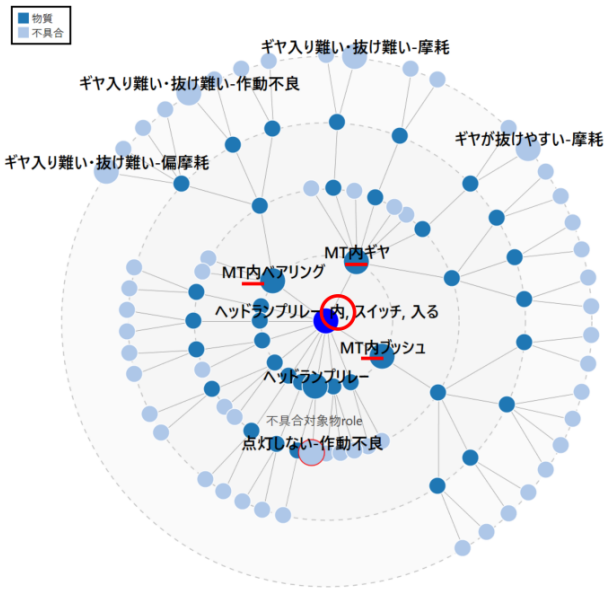


図 2: 灯火装置に関する機能ノードから
パワートレイン部に関する故障が出力された例

望ましいにもかかわらず，影響を及ぼさないはずの装置で故障が発生するという誤った推論結果が出力される問題が確認された．具体的には，図 2 に示すように，「ヘッドランプリレー内のスイッチが入る」という灯火装置の機能ノードを選択した際に,「ギヤが抜けやすい」や「ギヤが入り難い」といった，パワートレイン部に関する故障が出力される現象が見られた．

この現象の原因を調査したところ，故障推論を実行する際に選択した機能ノードを単語分割する過程で,「内」という単語を独立に抽出してしまい，その結果，オントロジー上で「内」を名称に含むような部品を対象に故障の影響を誤って伝搬させてしまっていた．図 2 は，実際の故障推論の出力結果である．パワートレイン部オントロジー内で定義されている「MT（マニュアル・トランスミッション）内ギヤ」や「MT 内ベアリング」といった「内」を名称に含む装置に，灯火装置の故障を誤って伝搬させてしまっていた．

パワートレイン部オントロジーでは,「内」という語を含んだ単語として，図 1 のように,「MT 内ギヤ」,「MT 内ベアリング」,「MT 内ブッシュ」,「AT（オートマチック・トランスミッション）内クラッチ」の 4 つが定義されていた．このうち,「MT 内ギヤ」は，MT の部品である各ギヤ（メインドライブギヤ，ローギヤ等）の総称として定義されている．また,「MT 内ベアリング」,「MT 内ブッシュ」も同様に MT の各ベアリング，各ブッシュの総称として用いられている．そして，これらの概念を用いて,「MT」の part of（p/o）スロットは，図 3 のようになっている．さらに,「AT 内クラッチ」についても，下位概念で「ハイクラッチ」,「フォワードクラッチ」といった AT 内で用いられているクラッチを定義しており，各クラッチの総称として用いられていることが分かる．このように，本来，p/o として定義すべき部品を，ラベルに含めて定義してしまっていたことが，問題であると考える．

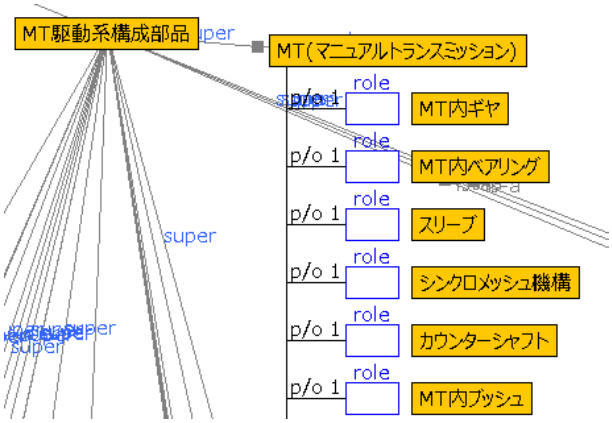


図 3: パワートレイン部オントロジーにおける
「MT」の定義（一部抜粋）

また，本研究で用いる機能分解木は，実現したい機能（上位ノード）を達成するために必要な手順を，下位ノードで記述している．例えば，図 6 に示すように,「MT で変則操作を行う」（上位ノード）の場合，3 速方式では図 6 のような構成となる．ここでは,「メインドライブギヤ」,「サードギヤ」といった具体的なギヤの名称を用いて，部分機能を表記している．この機能モデルに対して故障推論を行うと,「メインドライブギヤ」の機能に関する故障は出力される一方で,「MT」は図 3 のように定義されているため，機能分解木に「MT 内ギヤ」を主体とする機能を明記しない限り,「MT」に関する故障が推論されない．この点からも，定義の修正が必要であると考える．

4.2 知識モデル拡張時の問題点に対して実施した解決方法

前節で挙げた知識モデル拡張時の問題点に対して，本研究では，オントロジー上の定義方法の修正によって解決を図った．

「MT 内ギヤ」,「MT 内ベアリング」,「MT 内ブッシュ」,「AT 内クラッチ」という 4 単語は，いずれも「○○内△△」という形で定義されており，MT または AT で用いられている部品をそれらの下位概念で定義している．この場合，下位概念で定義した概念を「MT」または「AT」を特殊化した概念として定義することが可能である．具体例として,「MT 内ギヤ」に関して再定義したものを図 4 に示す．図 1 では，AT，MT，またはその両方に存在するギヤに関して，それぞれを「AT 駆動系構成部品」,「MT 駆動系構成部品」,「共通駆動系構成部品」の下位概念で定義していたが，図 4 では，AT，MT の部品にかかわらず「ギヤ」の下位概念です

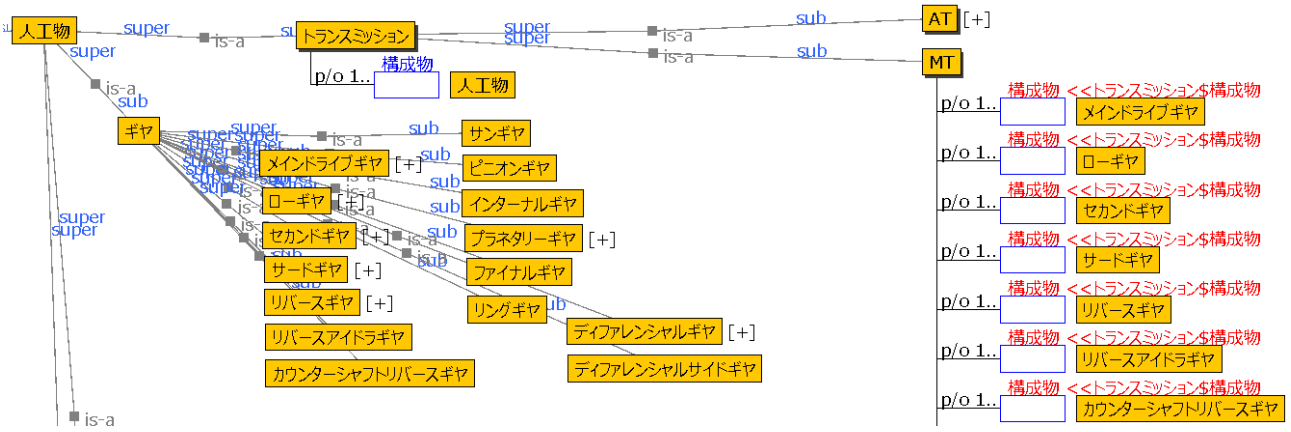


図 4: パワートレイン部オントロジーにおける「MT 内ギヤ」の定義修正

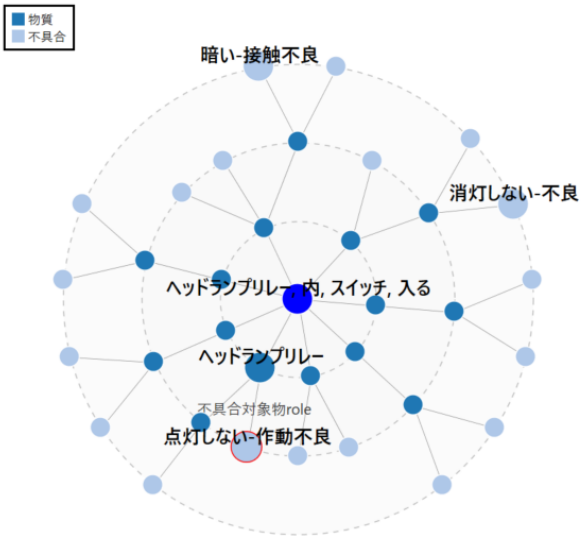


図 5: 修正後のパワートレイン部および灯火装置オントロジーにおける出力結果

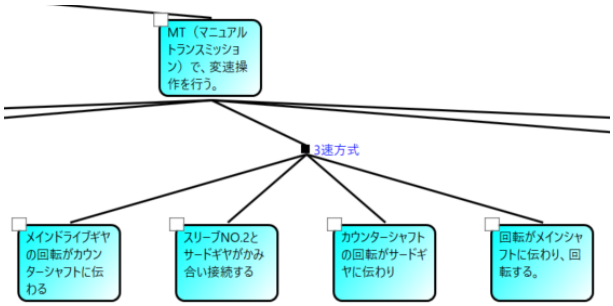


図 6: 自動車のパワートレイン部における機能分解木（一部抜粋）

べてを定義している．また,「トランスミッション」の p/o でロール概念「構成物」を定義し，その下位概念で定義した「MT」の p/o において，特殊化をすることで，MT の構成部品を定義している．

以上のように定義することで，本来，装置の内部を構成する部品群を,「内」という名称だけで定義していたために，間違った推論をしてしまっていたのを，装置を構成する部品であると p/o を使って定義したので，正しく推論できるようになった．

4.3 修正したオントロジーにおける故障推論の実行と考察

前節の「MT 内ギヤ」と同様に,「MT 内ベアリング」,「MT 内ブッシュ」,「AT 内クラッチ」に対しても，定義の修正を行った．その後，修正後のパワートレイン部オントロジーおよび灯火装置オントロジーを用いて，再度故障推論を実行した．その結果，図 5 に示すような出力結果が得られた．

図 2 で示した,「内」という語によりパワートレイン部の故障が誤って推論されてしまう問題は，図 5 において解消されており，灯火装置に関する故障のみが正しく推論されるようになった．よって，スロットや特殊化を用いて正しく定義することで，出力したい故障を推論できるようになることが分かった．したがって，オントロジー構築の際には,「内」といった語をラベルに含めるのではなく，part of という構造として表現する必要がある．

5 故障推論における自動車故障過程オントロジーの構築

自動車における故障推論を行うためには，故障に関する概念を体系的に定義し，オントロジー上で整理することが重要であると考える．しかし，先行研究 [3] のオントロジーでは,「故障」という概念自体の定義や分類が十分に行われておらず，故障がどのように発生し，どのように伝搬していくかといった故障の詳細な過程について表現できないものもあった．

本研究では，この課題を解決するために，来村・溝口による故障オントロジー [9] を参考とし，より具体的な故障の発生過程を捉えるための「自動車故障過程オ

ントロジー」を構築した．本オントロジーでは，自動車における不具合を「事象」の 2 つの下位概念「故障事象」と「伝搬事象」として定義する.「事象」は,「原因」,「結果」,「個所」,「時間」の 4 つのスロットで表現する.「故障事象」は不可逆な状態変化を伴う事象であり,「伝搬事象」は，内部状態が変化せず，異常入力により異常出力を行う事象である．

また，図 7 に示すように,「原因」は,「故障事象」の原因となる「故障原因」と,「伝搬事象」の原因となる「異常原因」に分類した．そして，それぞれの下位概念で，さらに前の段階に原因がある「相対原因」と, それ以上さかのぼれない最初の原因である「絶対原因」を定義した．「結果」についても，故障事象における結果の状態を表す「故障状態」と，伝搬事象における結果状態を表す「異常状態」を定義した．さらに,「時間」は，絶対異常原因より上流の因果連鎖にかかる時間を表す「故障時間」，および下流の因果連鎖にかかる時間を表す「故障後時間」を定義した．

故障過程の概念定義により，先行研究 [3] では一様に「不具合原因」とされていた関係を，事象の性質や因果的階層に基づいて分類することが可能となった．その典型的な事例として，灯火装置における「点灯しても暗い」という事象を挙げる．本事象の原因には，長期間の使用による「経年劣化」と，それに起因して発生する電球内部の「黒化現象」が存在する．本研究では,「経年劣化」をそれ以上さかのぼることのできない根本的な原因として「絶対故障原因」に分類し,「経年劣化」によって引き起こされる「黒化現象」を,「相対故障原因」として分類した．さらに，これらの原因により生じる「点灯しても暗い」という事象は，電球内部の物理的変化を伴う不可逆な現象であることから,「故障事象」として定義した．このように，原因の階層構造および事象の性質を明確にすることで，先行研究 [3] では一括して「不具合原因」として扱われていた要素を，因果関係に基づいて整理できることを確認した．

さらに，本研究では，図 8，9 に示すように，この自動車故障過程オントロジーを先行研究 [3] の灯火装置オントロジーに統合し,「ライトが点灯しない」「消灯しない」「点灯しても暗い」「左右とも点滅しない」「点滅回数が左右とも正しくない」「左右いずれかのランプが点灯したままになる」「左右いずれかのランプの点滅回数が正しくない」といった事象を故障事象または伝搬事象として分類・定義した．また，オントロジー上で関係概念を用いて，原因間の因果関係および時間的前後関係を明示的に表現した．これにより，ある事象に対してどの原因が根本的な「絶対原因」であり，どの原因がそれに続く「相対原因」であるかを，オントロジー上で視覚的に把握できるようになった．さらに,「絶対原因」の中でも「経年劣化」などの要因が発端となり，他の「絶対原因」が時間的に続いて発生するという関係をオントロジー上で可視化できるようになった．

故障過程の概念定義を精緻化することで，物質概念に 49 個，非物質の原因概念に 45 個，結果概念に 16 個，時間概念に 6 個，事象概念に 12 個，その他の非物質概念に 3 個の計 131 個の新たな概念を定義した．このような概念構築により，先行研究 [3] のオントロジーでは捉えきれなかった故障の多様な形態と因果構造を表現できる枠組みを構築したといえる．今後は，このオントロジーで故障推論が正しく実行できるかの検証を行い，その評価と改善を目指す．具体的には，自動車内の個別の系統内で発生した故障の影響が，正しく伝搬するか，伝搬しすぎないか，などを故障推論を行いながら評価したい．

6 結論

本研究では，機能を発揮する人工物において起こりうる故障シナリオをなるべく多く推論して導出する故障推論のスケール拡張と概念の精緻化を目的として，自動車のパワートレイン部および灯火装置を対象とした知識モデルの拡張と，故障に関しての概念の精緻化を行い，考察した．

今後は，引き続き，知識モデルの修正を行っていきたいと考えている．また，構築した「自動車故障過程オントロジー」について，実際に故障推論を実行し，出力結果として得られる故障シナリオの妥当性についても，検討していく予定である．

参考文献

[1] 飯田都楓，久保里紗，笹嶋宗彦，自動車設計業務効率化に向けた機能分解木とオントロジーによる故障知識を用いた仕組みの検討，人工知能学会全国大会論文集 第 38 回，pp.3Xin223-3Xin223，2024.

[2] 古川修，田村正隆，ダイナミック図解 自動車のしくみ パーフェクト事典 第 2 版，株式会社ナツメ社，2022.

[3] 島村佳周，飯田都楓，久保里紗，笹嶋宗彦，自動車の灯火装置を対象とした機能と故障のオントロジー構築と故障推論に関する基礎検討，人工知能学会全国大会論文集 第 39 回，pp.1L5OS1504-1L5OS1504，2025.

[4] 溝口理一郎，來村徳信，笹嶋宗彦，古崎晃司，オントロジー構築入門，株式会社 オーム社，2022.

[5] 古崎晃司，”法造 - オントロジーエディタ”，法造 - オントロジーエディタ 公開ページ，2023，https://www.hozo.jp/index_jp.html，(参照 2025-11-14)．

[6] 全国自動車整備専門学校協会，3 訂 自動車の故障と探求，株式会社 山海堂，2005.

[7] 來村徳信，笠井俊信，吉川真理子，高橋賢，古崎晃司，溝口理一郎，オントロジーに基づく機能的知識の体系的記述とその機能構造設計支援における利用，人工知能学会論文誌，17 巻，1 号 SP-C，pp.73-84，2002.

[8] 古崎晃司，來村徳信，溝口理一郎，生物規範工学オントロジーと Linked Data に基づくキーワード探索，人工知能学会論文誌,31 巻，1 号 LOD-D，pp.1-12，2016.

[9] 來村徳信，溝口理一郎，故障オントロジー: 概念抽出とその組織化，人工知能 14.5，828-837，1999.

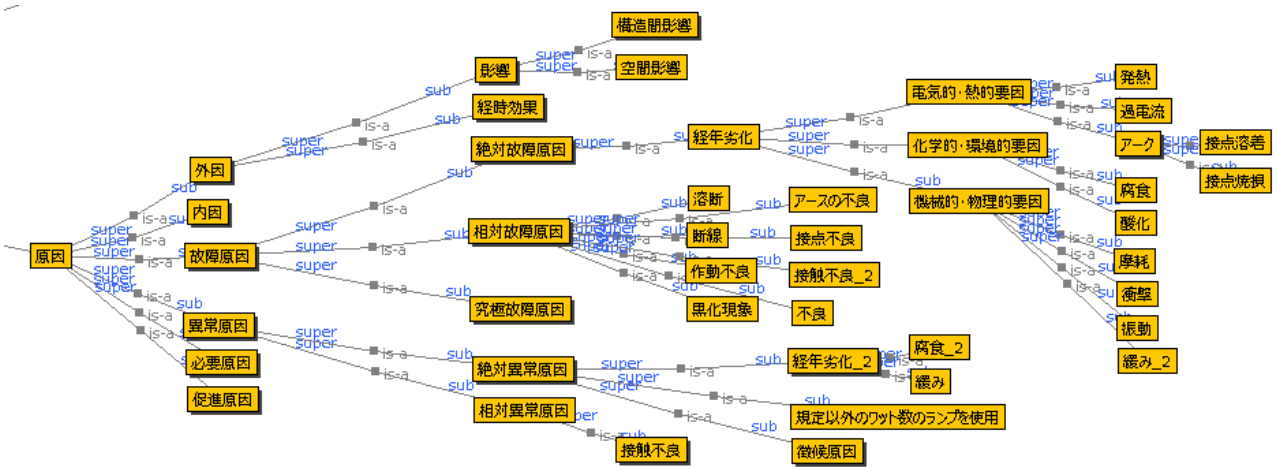


図 7: 來村ら [9] の故障概念オントロジーを基にした自動車灯火装置の故障概念の定義

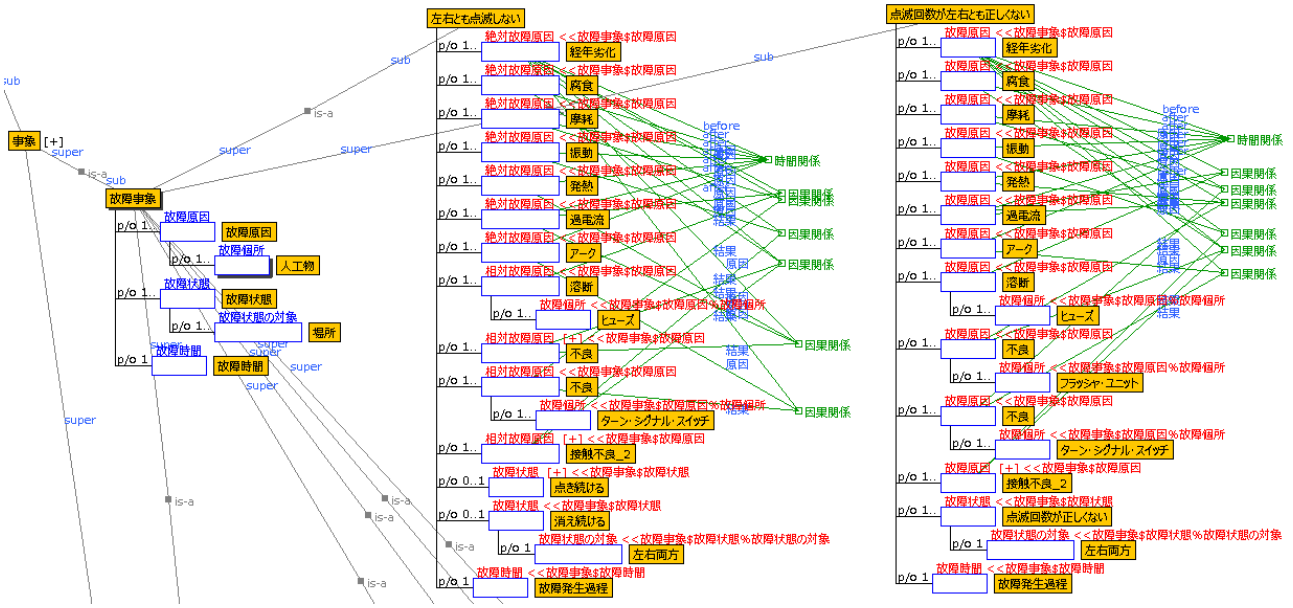


図 8: 先行研究 [3] と自動車故障過程オントロジーを統合して定義した「故障事象」(一部抜粋)

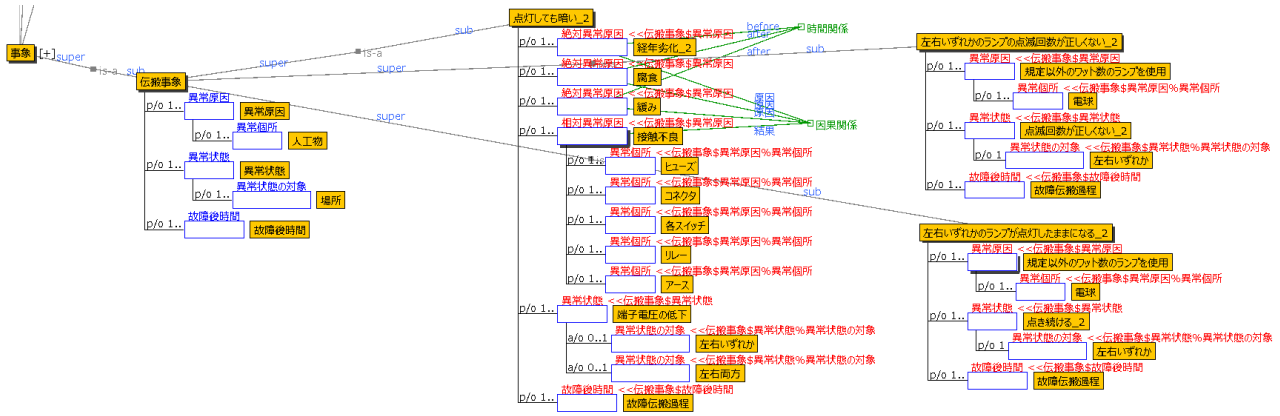


図 9: 先行研究 [3] と自動車故障過程オントロジーを統合して定義した「伝搬事象」

作業マニュアル作成初心者のための機能分解木構築方法論の検討

A Methodology for Constructing Functional Decomposition Trees for Beginners in Work Manual Creation

堺　貴彦[1]　　*笹嶋　宗彦[1]
Takahiko Sakai[1]　　Munehiko Sasajima[1]

[1] 兵庫県立大学
[1] University of Hyogo

Abstract: Functional decomposition trees are a useful method for structurally describing work manuals, including know-how, but they are difficult for beginners to create. This study analyzed functional decomposition trees created by beginners to identify common stumbling blocks and error patterns. Based on these results, we devised“Functional Decomposition Tree Construction Guidelines”and conducted an evaluation experiment. Experimental results revealed that determining the appropriate“level of detail for tasks”to create effective manuals is difficult for beginners. Furthermore, an evaluation experiment assessed the impact of the guidelines on beginners, confirming their effectiveness.

1　はじめに

機能分解木 [1] は，作業を目的と手段の関係に基づいて階層的に整理し，作業の構造や判断基準を明示的に記述できる表現手法である．この特徴により，作業手順だけでなく作業者の暗黙知や作業を行う際の意図を含めて，整理・共有することが可能であり，作業マニュアルの作成にも活用できるとされている [2].

筆者らの研究チームでは，協働ロボットを題材として，専門家が非専門家向けの作業マニュアルを自ら作成できるようにすることを目的とした研究を進めてきた．先行研究 [2] では，機能分解木を用いて，非専門家に作業手順を伝達するための「マニュアル構築ガイドライン」が検討され，その有効性が確認された．マニュアル構築ガイドラインは，非専門家に向けた，作業内容の説明方法や作業マニュアルを作成する上での注意点など，作業マニュアルの記述方針を定めたものである．

一方で，作業マニュアルの基盤となる機能分解木の構築方法については，具体的な手順や判断基準が十分に整理されておらず，多くの現場作業の専門家が機能分解木の初心者であるという実情に対応できていない．機能分解木は上位機能を達成するために必要な部分機能を階層的に整理し，その関係性を木構造として表現するものであり，作業の全体構造を把握するうえで有効であるが，初心者が適切な粒度や構造で記述することは容易ではない．

そこで本研究では，機能分解木の初学者を対象として，構築過程で生じる誤りやつまずきの傾向を分析し，それを踏まえて「機能分解木構築ガイドライン」を考案した [3]．さらに，ガイドラインの有効性を検証するため，2 回の評価実験を実施し，機能分解木の構築過程や作業粒度の改善に関する効果を検討した．その結果，作業の粒度に対する理解が深まり，利用者を意識した詳細な作業の記述が促される傾向が確認された．本論文では，第 1 回評価実験の結果とそれを踏まえたガイドライン改訂の内容，第 2 回評価実験の結果を報告する．

2　機能分解木について

本研究では，作業手順を階層的に記述する手法として機能分解木を用いた．本節では機能分解木について説明する．

機能分解木とは，実現したい機能を達成するために必要な部分機能を列挙し，それを木構造として表現したものである [1]．具体的には，上位機能を達成するために必要な部分機能を段階的に分解し，それぞれの関係性を明示的に示すことで，全体の機能構造を視覚的に整理する表現手法である．この木構造では，根の部分が最も上位の機能を表し，そこから枝分かれする形で部分機能を階層的に配置して表現する．

機能分解木におけるノードとは，部分機能として分解された特定の作業や機能を意味している．ノードに

*連絡先：兵庫県立大学社会情報科学部社会情報科学科
〒 651-2197 兵庫県神戸市西区学園西町 8 丁目 2-1
E-mail:ad24z024@guh.u-hyogo.ac.jp

は，作業や機能の目的，具体的な作業内容，作業や機能が達成される条件などが記載されることが一般的である．

また，方式とは，上位機能を実現するための具体的な手段や方法を指す [1]. 例えば,「木を切る」という機能を達成するためには,「のこぎりで切る」や「チェーンソーで切断する」などの異なる方式が考えられる．

適用条件とは,「作業を達成する上で満たしておくことが望ましい条件」とされている．阻害要因は,「その作業を行っている際に起きてはならないことやリスク」を記述し，その直下にリスクの解決策を記述するものである．例えば,「木を切る」という作業において，適用条件であれば「周囲の安全確認を行う」と記述し，阻害要因であれば「チェーンソーから異音がする」というリスクに対し,「作業を中断し，離れた場所でチェーンソーの点検を行う」という解決策を記述することが考えられる．

機能分解木において，ノードに記述された作業は，左から右へと進行し，その進行途中のノードに更なる下位ノードが存在する場合は，それらの作業を実施することで上位の作業が達成される．

3 初学者の作成した機能分解木の観察と考察

本研究では，2 名の機能分解木初学者が作成した機能分解木を対象に，機能分解木として適切に記述できているのかを基準として観察した．

なお，本研究における「機能分解木の初学者」とは，ノードや方式といった単語は知っているが，実際に自身で機能分解木を作成した経験がない人を指す．本節では，初学者の作成した機能分解木について観察と考察を行った結果の概要を示す．詳しくは [3] を参照されたい．

3.1 学生の作成した機能分解木

学生は，簡易症状問診システムの構築のために，病院における簡易症状診断を対象として機能分解木を作成した．簡易症状診断とは，病院の受付で発熱の有無などの病状に関して質問を行い，患者の緊急性や適切な診療科を判断するための初期情報収集作業である．学生が作成した機能分解木では，作業手順が上から下に流れるフローチャートのような形式となっており，横並びのノードがすべて同一の作業内容を持つなど，下位ノードの達成によって上位ノードが達成されるという機能分解木の基本的な規則が守られていなかった．また，このテーマである「簡易症状診断」は，症状の有無を確認する性質上，場合分けが多く，機能分解木で表現するには不向きなテーマであると考えられた．

そこで筆者らの研究チームは，症状の確認と患者への指示を行う作業が機能分解木で適切に記述できるかを検証することを目的として，新たに機能分解木を作成した．その結果，重複していた場合分けを統合することで，患者への指示が決定した時点で診断を終了し，指示を行うことができる構造となった．このことから，場合分けの多いテーマであっても，必要な場合分けと不必要な場合分けを整理し，必要最小限に抑えることで，機能分解木として適切に表現できることが明らかとなった．

3.2 作業の専門家が作成した機能分解木

溶接ロボットのユーザーの基礎教育を目的として，作業の専門家は「溶接コアトレーニング」を対象に機能分解木を作成した．溶接コアトレーニングとは，非専門家が溶接作業を安全かつ効率的に行うために必要な知識やスキルを学ぶ内容である．専門家は，作業の粒度を適切に設定することが最も難しく，特に想定するマニュアル利用者のスキルレベルに応じて調整が必要である点を課題として挙げた．また，適用条件（黄色ノード）は「できれば満たしておきたい条件」，阻害要因（ピンクノード）は「絶対に満たさなければならない条件」として使い分けるなど，独自の運用が見られた．さらに，方式ノードの記述ルールが不明瞭であったため，どのようにノードを分けるべきか判断に迷う場面があったと述べており，参考となる機能分解木のサンプルを提示することで，これらの課題が解決できると考えられる．

4 初学者に向けた機能分解木構築ガイドラインの検討

本研究では，初学者による機能分解木の作成で確認された課題を踏まえ,「機能分解木構築ガイドライン」を検討した [3]．このガイドラインは，初学者が疑問に感じやすい問題について記載しておくことで，適切な形式の機能分解木を作成できるようになることを目的としている．本節ではその内容について説明する．

機能分解木構築ガイドライン案：

1. 作業内容の理解に必要なノード数は作業の難易度によって変化するため，ノード (作業手順) 数は必ずしも少なければ良いというわけではない．

2. 初学者が行うことを想定して可能な限り具体的に作業の (機能) 分解を行う．主観的な判断で記述を省略したり粒度を大きくしたりすることは避けるべきである．(マニュアルを利用する際は自身のスキルレベルに応じて，表示する作業の抽象度や粒度を調整する)

3. 作業は左から右方向へ順になるように書かなければならない.
4. 下位ノードが全て達成されることで上位ノードが達成されなければならない.
5. 1 つのノードにつき作業は 1 つしか入れてはならない.(手順が複合してはならない)
6. 作業を達成する上で用いる方法 (原理) が異なる場合は「方式」を分ける.
7. 適用条件は 1 つ右の作業を達成する上で満たしておくことが望ましい条件として記述する.また,阻害要因はリスクとして「失敗の内容」を記述し、その直下に青色ノードで「解決策」を記述する.

ガイドライン第 2 項目のノードと粒度について述べる.ノード数については,「ノード（作業手順）数は必ずしも少ないほど良いというわけではない」とし,作業の難易度に応じた柔軟な調整が必要であると考えた.作業の難易度が高い場合には,より詳細な作業手順を示すために下位ノードを増やすことが適切であると考えた.

また,マニュアルを利用する非専門家のスキルレベルに応じて,作業手順の詳細度を調整する必要性についても検討した.専門家が作成した機能分解木では,前工程で行った作業は理解しているものとして,記述を省略することが見られた.省略された部分の作業を見ると,その記述のみでは作業の達成が難しく,前工程まで電子マニュアル上で遡り,現在行っている作業のノードと往復しながら確認する必要があった.これは結果として作業効率を悪化させるため,作業に関する記述は可能な限り詳細に行うべきであると考えられる.この点に関しては「想定される利用者の知識レベルに応じて作業について詳細な記述を省略しがちであるが,作業マニュアルとして利用するためには可能な限り詳細な記述を心がけるべきである」と考えた.

一方で,作業を詳細に記述する場合,ノード数が増加しマニュアルが膨大になってしまうという問題に関しては,初心者が下位ノードに詳細な手順を記載した内容を通じて,正確に作業を達成できるようにすると同時に,熟練者は上位ノードを中心に閲覧し,必要に応じて下位ノードを展開して読むという形式を検討した.この形式は,機能分解木が抽象度に応じて階層化されているという性質と,FWT エディタのノードをクリックすることで下位ノードを展開,収納して表示できるというインタフェース機能を利用したものである.初心者は,葉のレベルまで表示したマニュアルを参照することで詳細に作業を理解できる一方,熟練者は,下位の階層を収納して上位階層のノードのみを表示し,必要最小限の情報のみを参照することで,迅速に作業を行うことが可能になると期待される.このことから,「マニュアルを利用する際は自身のスキルレベルに応じて,表示する作業の抽象度や粒度を調整する」という内容をガイドラインに取り入れた.すべての作業について網羅的に詳細な記述を行うと,電子マニュアルでの工程数が膨大となるため,現場で実際に使用する場合に,読み飛ばしのエラーが起きる場合があるなどの問題が先行研究 [2] で報告されているが,この形式を用いることで,初心者と熟練者の両方のニーズに対応した作業マニュアルとして活用できるのではないかと考えられる.

また,機能分解木における基本原則である,「下位ノードが達成されることで上位ノードが達成なければならない」や「1 つのノードにつき作業は 1 つしか入れてはならない (手順が複合してはならない)」について明記した.これらは基礎的な内容ではあるが,機能分解木の構築において不可欠な要素であり,その重要性からガイドラインに取り入れるべきであると判断した.また,「作業は左から右方向へ順になるように書かなければならない」という項目については,初学者が作成した機能分解木において,上から下方向に作業が進行するフローチャート型の機能分解木が作成されたことを受けて,ガイドラインの一部として取り入れた.

さらに,方式をどのような判断基準で用いればよいのかについては,機能分解木を作成した際に,初学者から「方式の使い方がよくわからなかった」という指摘を受けた事を踏まえ,「作業を達成する上で用いる方法 (原理) が異なる場合は方式を分ける」という内容をガイドライン案に取り入れた.この指針により,作業を達成する上で用いる道具や方法,原理が異なる場合には,方式を分けて記述しなければならない,ということについて, 初学者が理解しやすくなると考えた.

最後に,適用条件と阻害要因に関しては,初学者が独自の解釈で利用していたことから,適用条件は「右の作業を達成する上で満たしておくことが望ましい条件」であり,阻害要因は,リスクとして「失敗の内容」を記述し、その下に青色ノードで「解決策」を記述する,という内容を新たなガイドライン項目として取り入れた.これにより初学者が適用条件と阻害要因を正しい形で利用できるのではないかと考えられる.

5 提案ガイドライン評価実験 1

第 1 回評価実験では,学生 10 名に「自身の得意な作業」をテーマとして,機能分解木を作成してもらった.この際に,機能分解木構築ガイドラインを配布し,作成される機能分解木にどのような影響を与えるのかを観察した.本節では,第 1 回評価実験の結果の考察と得られた知見から修正を行った「機能分解木構築ガイ

ドライン第 2 版」について説明する．

5.1 実験概要

学生 10 名に対し，アルバイトの業務内容や趣味などの「自身の得意な作業」について，その作業を行ったことのない初心者に作業内容が伝わるような機能分解木を作成するというテーマを与えた．まず，機能分解木の概要および機能分解木を作成するためのツールである FWT エディタの操作方法について説明を行った．その後，各学生に，テーマに沿った機能分解木の作成を行わせた．作成作業終了後に,「機能分解木構築ガイドライン」を学生に配布し，ガイドラインの各項目の内容について解説を行った．自身の作成した機能分解木の中で，誤った記述をしていた箇所や表現できていなかった作業について，配布したガイドラインを参考に，修正・追記作業を行わせた．その後，学生に「ガイドラインがあることで機能分解木は作成しやすくなったのか」,「ガイドラインのどのような点が機能分解木作成において役立ったのか」,「機能分解木作成においてどのような要素の理解が難しいと感じたか」についてのアンケートを実施した．実験の流れを以下に示す．

1. 機能分解木と FWT エディタの基本的な機能について説明を行う．
2. 機能分解木構築ガイドラインなしで,「自身の得意な作業（趣味・アルバイトなど)」についての機能分解木を作成してもらう．
3. 機能分解木構築ガイドラインを配布し，内容について説明を行う．
4. ガイドラインを参考にして，2. で作成した機能分解木について修正・改良を行ってもらう．
5. ガイドラインについてのアンケートに回答してもらう．

5.2 実験結果

学生 10 名が作成した機能分解木の分析結果およびアンケートの結果について示す．得られた結果は，作成された機能分解木の内容分析とアンケート回答の 2 つの観点から整理した．

学生が作成した機能分解木の内容を分析した結果，多くの学生が「1 ノードにつき作業は 1 つしか入れてはならない」という原則を正しく適用しており，また「方式」の利用についても概ね適切に記述できていた．さらに,「適用条件」や「阻害要因」といった機能分解木特有の記述方法についても，自身のノウハウをマニュアル利用者へ伝えるために活用できている例が見られた．これらの点から，作業マニュアルとして用いる上での機能分解木の正しい記述方法の理解に，ガイドラインの提示が一定の効果をもたらしたと考えられる．

一方で,「作業の粒度（どの程度まで作業を詳細に記述するか)」については，5 名の学生の作業内容の記述が不十分であり，このままでは，マニュアルの利用者に伝わらないと考えられる箇所が多数見られた．この点は，ガイドラインに示す指針や，配布資料に記載する具体例の内容に，改善の余地があると考えられる．

アンケート結果においても同様の傾向が見られた．アンケート結果を図 1 に示す．

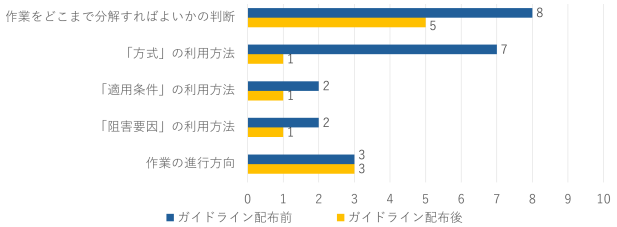


図 1: 機能分解木を (ガイドラインなし/あり) で作成した際に，難しいと感じた点や理解できなかった点にすべてチェックを入れてくださいに対する回答

「方式」の利用方法（第 6 項目）については，図 1 の選択式回答で，ガイドライン配布前には 7 名が「難しい・理解できなかった」と回答していたが，ガイドライン配布後には 1 名まで減少した．また,「1 つのノードに作業は 1 つしか入れてはならない」（第 5 項目）については,「複数の作業を 1 つのノード内に記述していたが，ガイドラインを参考にすることで間違いに気付くことができた」といった回答が記述式回答から得られた．以上のことから，ガイドラインが初学者の理解度向上に有効であることが示唆された．特に，配布資料中で正解例と誤例を比較して示したことが，理解度の向上に寄与したと考えられる．

一方,「作業の粒度」および「作業の進行方向」に関しては依然として理解が不十分であり，後者については，ガイドライン配布前後で大きな変化が見られなかった．

これらの結果を踏まえ,「機能分解木構築ガイドライン第 2 版」として「作業の粒度」に関する新たな指針を検討する．また,「作業の進行方向」については，より具体的な解説を配布資料に追記し，作業の進行順序を視覚的に示すことで理解の支援を行った．

5.3 機能分解木構築ガイドライン第 2 版

第 1 回評価実験の結果を踏まえ，機能分解木構築ガイドラインの内容を見直し,「機能分解木構築ガイドライン第 2 版」を作成した．機能分解木構築ガイドライン第 2 版では，主に「作業の粒度」に関する指針の明確化と,「作業の進行方向」に関する配布資料の改良を行った．

まず,「作業の粒度」(第 2 項目) については,「作業の粒度 (どこまで詳細に説明するか) はマニュアルの試作と実験を繰り返す過程において，マニュアル作成者と利用者との間で合意を形成することで決定するべきである.」という指針へと改訂した．これは，ガイドライン第 1 版の第 2 項目における「初学者が行うことを想定して可能な限り具体的に作業（機能）の分解を行う．主観的な判断で記述を省略したり粒度を大きくしたりすることは避けるべきである．マニュアルを利用する際は，自身のスキルレベルに応じて，表示する作業の抽象度や粒度を調整する.」という内容を見直したものである．第 1 回評価実験の結果から，初学者にとって「可能な限り詳細に作業を記述する」という指針は抽象的であり，結果として作業内容の記述が不十分な機能分解木が多く作成された．

この結果を受け，ガイドライン第 2 版では，マニュアルの作成を行う作業の専門家とマニュアル利用者の双方が，どの程度まで作業を詳細に記述すべきかを議論し，合意を形成する過程を経て，適切な粒度を決定するべきであるという指針を示した．

また，配布資料の改訂として,「作業の粒度」に関しては，機能分解木上で上位ノードから下位ノードに進むにつれて，作業の粒度が「抽象的」から「具体的」へと変化することを視覚的に理解できるように，実例を追加し，解説する項目を新たに設けた．図 2 に配布資料に追記した内容を示す．

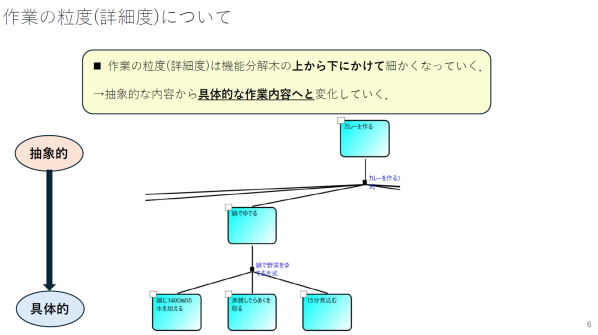


図 2: 作業の粒度

さらに,「作業の進行方向」については，単にノードに番号を付与するだけでなく，作業がどのような順序で進行していくのかを具体的に示す説明を追記した．配布資料に追記した内容を図 3 に示す．これにより，機能分解木における作業の流れをより明確に把握できるよう配慮した．

6 提案ガイドライン評価実験 2

第 2 回評価実験では，第 1 回評価実験で明らかになった課題を踏まえ，改訂したガイドライン第 2 版の有効性を検証することを目的とした．特に，第 1 回評価実験において初心者が困難を感じていた「作業の粒度」に着目し，ガイドライン第 2 版に新たに追加された「作業の粒度について合意を形成する過程」をフィードバックとして位置付け，この過程が「作業の粒度」にどのような影響を及ぼすかを観察した．本節では，その詳細について説明する．

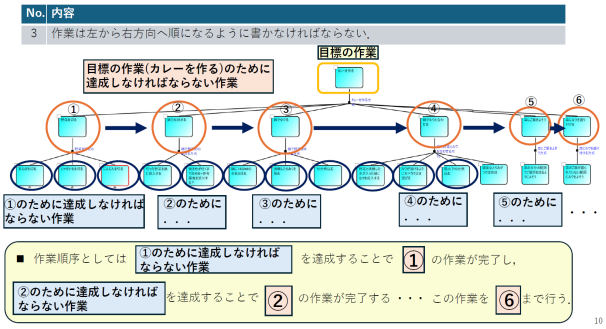


図 3: 作業の進行方向

6.1 実験概要

第 2 回評価実験は，学生 8 名を対象に実施した．参加者には，第 1 回評価実験と同様に「自身の得意な作業」をテーマとして機能分解木を作成してもらった．第 1 回評価実験との主な差異は, 機能分解木構築ガイドライン第 2 版に新たに追加した「作業の粒度 (どこまで詳細に説明するか) はマニュアルの試作と実験を繰り返す過程において，マニュアル作成者と利用者との間で合意を形成することで決定するべきである.」という指針に対し，作業の初心者であるマニュアル利用者の視点から，フィードバック資料を作成し，マニュアル作成者に配布するという過程を経ることで，初心者の作成する機能分解木にどのような影響を与えるのかを観察するという点である．実験の工程としては，第 1 回評価実験の内容に加えて，ガイドライン配布後に作成された機能分解木に対し，研究チーム内で議論を行い，フィードバック資料を作成者に配布し，それを参考に修正・追記作業を行うという工程を追加した．

第 1 回評価実験と同様に，1〜4 の手順を実施した後，以下の手順を追加で行った．

5. 提出された機能分解木に対して，実際に作業が実施可能かどうかを基準にフィードバック資料を作成する．
6. 参加者にフィードバック資料をもとに，再度，修正・改良を行ってもらう．
7. 参加者にアンケートに回答してもらう．

6.2 実験結果

作成された機能分解木を分析した結果，多くの学生が与えられたフィードバックを参考に修正作業を行い，後述する通り，マニュアルの想定利用者が求める作業の粒度について理解を深めた様子が確認された．学生がフィードバック資料を利用し，修正・追記作業を行った機能分解木を図 4 に示す．

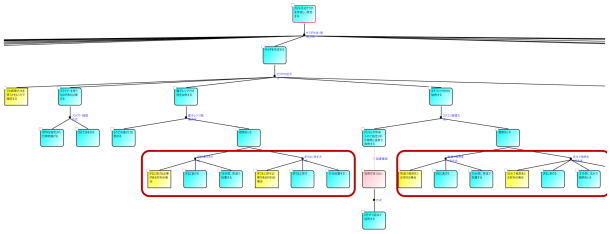


図 4: 学生が修正追記作業を行った機能分解木 (一部)

作業マニュアルの想定利用者を「対象とする作業の初心者」として意識し，作業をより詳細に記述しようとする傾向も見られた．たとえば，フィードバック資料において「粗熱をとる」という作業ノードに対し,「どのような作業を行うことで達成されるのかを具体的に記述してください」と指摘したところ，修正後の機能分解木では「ざるにあげる方式」と「ボウルに移す方式」の 2 種類が新たに作成され,「○分間常温で放置する」といった具体的な作業手順が追加されていた．また,「風袋を測定する」という記述に対しては,「風袋（パックのみの重さ）を測定する」と追記され，作業マニュアルとして用いる際に，ノード単体で意味が理解できるように改善されていた．このように，フィードバックを通じてマニュアル利用者が求める粒度へ近づける修正が行われていたことが確認された．

アンケート結果からは，フィードバック資料に関する意見や，新たな課題が見られた．フィードバック資料に関して，実験参加者からは,「機能分解が不十分な点を具体的に指摘してくれた点が修正・追記作業に役立った」,「修正の際に意識してほしい部分を太字にしていることで，何を意識して修正すればよいかが分かりやすかった」といった意見が記述式回答から得られた．これらの結果から，フィードバック資料の内容は修正・追記作業に適した形式であったと考えられる．一方で,「適用条件」に関しては，ガイドラインには「達成が望ましい作業」と記載していたが，通常ノードの「達成すべき作業」との明確な区別が難しく感じたという意見が得られた．また，新たな課題として,「作業の中で『してはならない作業（禁止事項)』について，どのように記述すればよいのかがわからなかった」といった意見も記述式回答から得られた．

6.3 考察・今後の検討

第 2 回評価実験を通して，ガイドライン第 2 版が「作業の粒度」の理解度向上に寄与する傾向が見られた．特に，フィードバックを通じて修正作業を行う過程が，マニュアル作成者にとって初心者の求める作業の粒度を理解する上で，有効な取り組みであることが示唆された．一方で,「適用条件」と「通常ノード」の違いについては，配布資料に，適切な事例を示しつつ，明確に説明する必要があることが明らかとなった．また,「してはならない作業」については，従来の機能分解木の枠組みでは直接的に表現することが難しいため，新たな記述方法の定義が必要と考えている．

7 おわりに

本研究では，機能分解木構築ガイドライン第 1 版を基に，第 1 回評価実験の結果から得られた知見を反映し，機能分解木構築ガイドライン第 2 版を提案した．ガイドラインの改善では，初学者が特に理解に苦しんでいた「作業の粒度」に関する指針を中心に見直し,「作業の進行方向」については，具体例の提示や図による補足を追加することで，理解を支援する工夫を行った．第 2 回評価実験の結果，ガイドライン第 2 版の提示によって，作業の粒度に対する理解が深まり，利用者を意識した詳細な記述が促される傾向が確認された．さらに，フィードバックを通じた修正作業が，マニュアル作成者にとって利用者の視点を理解し，より適切な粒度を形成するうえで効果的であることが示唆された．

今後は,「してはならない作業（禁止事項)」などの，現在の機能分解木の記述方法では表現の難しい事象に関して，作業マニュアルとして適した記述方法を検討するとともに，初学者にとってより有用なガイドラインへと発展させることを目指したい．

参考文献

[1] 來村徳信, 他, オントロジーに基づく機能的知識の体系的記述とその機能構造設計支援における利用, 人工知能学会論文誌, 17 巻, 1 号, 2002.

[2] 平岡 あおい，山口 知彦，笹嶋 宗彦，2 機種の協働ロボット導入マニュアルへの機能分解木技術の適用と非専門家向けマニュアル作成方法論の検討，人工知能学会合同研究会 2024 インタラクティブ情報アクセスと可視化マイニング予稿集，SIG-AM-33-01，pp.1-8，2024.

[3] 堺 貴彦，平岡 あおい，山口 知彦，笹嶋 宗彦，マニュアル作成のための初学者に向けた機能分解木構築方法論の一検討，第 39 回人工知能学会全国大会論文集，1L5-OS-15-03,2025.

メタデータおよびユーザ行動に基づくマルチビュー融合による データセット間類似度の学習手法

程昊陽 [1*]　　早矢仕晃章 [1]
Haoyang Cheng[1]　Teruaki Hayashi[1]

[1] 東京大学 大学院工学系研究科
[1] School of Engineering, The University of Tokyo

Abstract: A dataset's metadata, consisting of structured descriptors such as titles, tags, and descriptions, summarizes the dataset's topic, provenance, and structure without requiring access to its content. We propose a multi-view framework for learning dataset–dataset similarity solely from metadata. The available information is partitioned into two metadata views—Tag and Text—and an auxiliary non-metadata view, User Behavior. For the Tag and Text views, we construct Dataset–Tag/Word bipartite graphs, perform type-constrained random walks, treat the walks as sentences, and train Skip-gram with Negative Sampling (SGNS) to capture contextual co-occurrence, from which we derive per-view dataset–dataset similarities. For the User Behavior view, we model co-usage sequences using Item2Vec to obtain an auxiliary similarity. Finally, the three views are integrated through Adaptive Fusion to produce a unified similarity matrix over datasets. Experiments on the MetaKaggle dataset demonstrate that the proposed method outperforms standard baselines in a series of metrics including nDCG@20 and MAP@20.

1　はじめに

異なる分野のデータを組み合わせた価値創出がイノベーションの源泉として注目されてきている．このような中，Web 上のプラットフォームにおいてデータ提供者がデータまたはデータに関する情報を公開し，利用者がデータを検索・購入するデータ市場が発展してきた．近年の調査では，このようなデータ市場やデータ取引プラットフォームが世界的に登場してきており，金融，都市ガバナンス，医療など幅広い分野において重要インフラとして位置づけられはじめている [1].

活用可能なデータの種類や規模が増大する一方で，異種のデータの探索から発見，活用までのプロセスの随所で様々な問題が表出してきている．例えば，個々のデータセットはしばしば構造が複雑でサイズも大きく，加えてプライバシー規制やコンプライアンス要件を伴う場合も少なくない．そのため，データを実際にダウンロードして内容を確認し，前処理を行ったうえで類似性を判断するには多大なコストがかかる [2]．既存のデータセット検索システムやデータレイク内の探索手法は，タイトルや簡単なキーワード，表層的な特徴量に依存しており，異種性やノイズを含む環境では自身の関心に合致したデータセットを見つけることは困難である．結果として，データ再利用やモデル転移などの応用の妨げとなっている [3].

本研究ではこれらの課題に対し，マルチビュー融合によってデータセット間の類似度を推定する手法を提案する．まず，タグ情報，テキスト情報，ユーザ行動といった異なるデータに関する情報源をビュー（View）として扱う．そして，Tag/Text ビューにおいて二部グラフ上の型制約付きランダムウォークによって得られる系列を文と見なし，Skip-gram with Negative Sampling（SGNS）を用いて表現ベクトルを学習する．さらに，近似最近傍探索（Approximate Nearest Neighbor, ANN）や正規化手法を用いて各ビューの類似度行列を構築する．これら複数の類似度行列は，マルチビュー融合によって統合し，最終的に単一の類似度行列を得る．

提案手法は，従来の実データに基づく手法や単一ビューによる手法と比較して，大規模な実データへのアクセス制限や異なるデータ形式による制約を回避しつつ，タグ，テキスト情報，ユーザ行動という異なる情報源の相補的な特性を統合することができる．その結果，単一の情報源に基づくアプローチでは捉えにくかった，高次の意味的関係や潜在的な利用パターンを抽出でき，より有効なデータセット類似度の推定が可能となる．

*連絡先：東京大学大学院工学系研究科
〒 113-8656 東京都文京区本郷 7-3-1
E-mail: teikoyo@g.ecc.u-tokyo.ac.jp

2 先行研究

2.1 実データによる表形式データ類似度推定

データの内容，すなわち実データに基づく表形式データの類似性推定やデータ探索・発見研究は，セルの値や統計的特徴を利用して表データ同士の類似度を計算し，検索やマッチング，統合タスクに適用されてきた．例えば，Lv らは Ferret を提案し，特徴量が豊富なオブジェクトを対象として，大規模データから内容が類似するオブジェクトを探索する枠組みを提供した．しかし，構造化した表や意味情報の扱いには制約があるという課題があった [4]．また，Yan らは Web クエリに対して実データベースのテーブル検索手法を提案し，表形式データの内容に基づくランク付けによりマッチング精度を改善した．だが，この手法は主に単一クエリによる検索に焦点を当てており，表全体の類似関係を体系的に表現する手法ではない点に課題がある [5]．作本らは，多面的なデータセット分類基準を設定し，類似データセット発見タスクにおいて，様々な類似度指標の有効性を評価した．しかし，この研究の主な関心は指標選択と評価であり，個々のデータのどのような情報が類似度に寄与しているのかということや計算コストの問題には十分に踏み込んでいない [6]．

2.2 メタデータによるデータ類似度推定

メタデータに基づくデータの類似度推定では，タイトル，説明文，タグ，変数情報などの補助的属性を利用してデータセット間の関連性を評価することで，異種のデータ検索や統合を支援する方法が提案されている．例えば，Ravishankar らは，少数のメタデータ属性のみを用いた教師なしクラスタリング手法を提案し，実データへアクセスせずにクラスタリング品質を向上させようとした．しかし，主な対象は一般文書やレコードであり，データセットレベルでの類似性推定は限定的であった [7]．Bernhauer らは，文脈情報を組み込んだオープンデータの類似検索フレームワークを提案し，関連データセットの発見率が高まったと報告している．しかし，メタデータの疎密やノイズに対して脆弱であるという課題が残っている [8]．また，Sakumoto らは，メタデータによるクラスタリングとメタデータ項目選択手法を提案し，複数分野において異なるメタデータ類似度指標とその組み合わせを体系的に比較した．これにより類似データセット発見やデータ連携に有用な知見を提供したが，依然としてメタデータ内部における指標選択や重み付けが中心であり，異種メタデータ間の補完関係を統合的に活用する枠組みの提案には至っていない [9]．

2.3 マルチビュー融合とグラフ類似度の融合

マルチビュー融合とグラフ類似度の統合とは，相補的な複数の情報源であるビューが存在する場合に，それらを統合して一貫した類似度表現を構築し，単一ビューよりも安定したクラスタリングや分類性能を得る方法である．Kumar らは共正則化マルチビュースペクトラルクラスタリングを提案し，各ビューのスペクトル埋め込み間に一貫性のある正則化項を入れることでクラスタリング精度を向上させた．しかし，この手法はビュー間の品質差やノイズビューの影響を自動的に抑制する仕組みが十分ではない [10]．Wang らの Similarity Network Fusion は，各ビューをサンプル類似度グラフとして構築し，反復的なグラフ拡散により統合する手法であるが，ビューごとの類似度行列を事前に固定して与える必要があり，拡散パラメータやビュー重みの設定に敏感であるという課題がある [11]．さらに，Gönen らは複数ビューのカーネルに局所的な重みを付与することでサンプルの局所構造を捉え，クラスタリング性能を改善した．しかし，この手法は事前に設計または計算されたカーネルに依存するため，ビューの種類や構造が大きく異なる場合には直接適用しにくい [12]．

これらの手法はマルチビューとグラフの融合の有効性を示しているものの，ビュー間類似度をあらかじめ適切に構成し，ビュー同士も比較的同質であると仮定することが多い．そのため，出自や構造が大きく異なる複数のメタデータからなるビューや，その品質のばらつきや不均衡を細かく一貫して扱うモデルを統合したフレームワークは十分に整っていない．

3 提案手法

3.1 メタデータのマルチビュー表現

メタデータとは，データセットに関する情報を形式的に記述したものであり，データ名，概要，タグや利用履歴などの情報を含んでいる．メタデータは分析対象そのものではなく，主にデータセットの説明書として機能し，検索に用いられることが多い．実データの扱いと比較し，メタデータを使うことにはいくつかの利点がある．例えば，メタデータは軽量かつ少量の記述のみで大まかなデータセットの内容や関連性を判断するための情報が含まれている．そのため，アクセス性や計算のコストが低く，プライバシーやコンプライアンス上のリスクが小さい．また，異なる種類のデータセットでも，共通の記述項目によって横断的にデータ同士を比較することができる．さらに索引構築・クラスタリング・推薦などの基盤をメタデータ層で整備することができる．そのため，本研究では実データで

はなく，メタデータを対象としてビューを構築するアプローチを採用する．

本研究のビュー（View）とは，データセットに付帯するメタデータ群から抽出された特定の特徴空間を指す．言い換えれば,「ある観点から観測されたデータセットに関する情報」をである．また，実験ではデータセットに Meta Kaggle データセット（4.1 節にて後述）を用いるため，以降のビューの作成については Meta Kaggle データセットの構造をもとに説明する．まず，本研究ではメタデータの特徴や性質を踏まえ，メタデータをタグビュー，テキストビュー，行動ビューの 3 種類に分けて用いる．

タグビューはメタデータの `Tags` の項目を利用する．Tags には，データセットの主たる分野，データ形式，代表的な利用シーンが記述されている．まず，タグ文字列を小文字化し，空白の削除・分割などで正規化する．そして，全データセットでの出現頻度から代表的なタグ集合を抽出する．同一または近しい概念を表すタグ（computer science, tabular, image など）を共有するデータセットペアは，研究領域やデータタイプ，データの用途が近いと見なすことができる．タグビューは「大まかに同種のデータセットか」ということを把握するための情報となるが，概念の粒度が統一されていなく，粗いものもあり，データ提供者の主観的に左右されやすい．

テキストビューは `Title` や `Description` などの自然言語による自由記述の項目で構成される．これらはタグより詳細な情報を有し，内容・構造・利用方法などが記述される．本手法ではこれらを 1 文書として結合し，前処理と分かち書きを行った後，テキストビューの特徴空間を構築する．テキストビューは「内容が似ているか」,「同じタスクに利用できるか」といった詳細な意味や内容の類似性が計算でき，タグビューの不十分さを補完する一方で, 長文・ノイズ・冗長記述も多いため，他のビューとの併用が重要となる．

ユーザ行動ビューは，作成者・組織 ID，閲覧数，ダウンロード数，投票数，利用頻度，利用履歴の時系列などで構成される．同一ユーザによって作成・管理されているデータセットや，利用パターンが類似するデータセットは,「利用方法が似ている」と見なすことができる．そのため，このビューは「誰がどのように当該データセットを利用しているのか」というユーザの類似性を捉え，データセットの内容や他のビューでは得られない機能的・利用文脈的な近さを反映することができる．

3.2 タグ・テキスト二部グラフの構築

3.2.1 タグビュー（Tag View）の二部グラフ構築

タグビューでは，まずデータセット集合（D）とタグ集合（T）から二部グラフ D–T とを生成する．はじめに，タグ文字列に対して小文字化，空白の除去，分割などの前処理を行う．続いて，全データセットにおける出現頻度（w_{tag}）を計算し，10 件以上のデータセットに出現するタグのみを残す．そして，各データセットとそこに含まれるタグの組を三つ組 $(d, t, 1)$ として取り出す．ここで 3 番目の要素 「1」 は,「そのタグが当該データセットのメタデータに現れた場合，対応する辺の重みを 1 とする」ことを表す．これらの三つ組を集約して，$|D| \times |T|$ の疎行列 D–T を得る．

3.2.2 テキストビュー（Text View）の二部グラフ構築

テキストビューでは，はじめに，テキスト形式で提供されているメタデータの項目である `Title`，`Subtitle`，`Description` を連結し，各データセットのテキストのリストを作成する．続いて，小文字化や正規表現を用いた前処理を行った後，トークン列の生成を行う．そして，トークンの文書頻度を算出したうえで，少なくとも 200 データセットに出現し，全データセットの 50% 以下にしか出現しない語のみを残す．最終的に，データ（D）とトークン（W），その出現頻度（w_{word}）の 3 つ組を得，$|D| \times |W|$ の疎行列 D–W を構築する．これにより，テキスト情報を自然言語由来の意味情報に符号化し，タグビューと並列に扱えるグラフの構造であるテキストビューに変換する．

3.3 ランダムウォークによる文生成

本節では，タグビュー（D–T）およびテキストビュー（D–W）に対して，データセット間の高次の共起構造を抽出するためのシーケンス生成手法を説明する．この目的は，二部グラフに埋め込まれたタグと語を介した間接的な関係性をシーケンスとして取り出し，SGNS による表現学習のコーパスとして活用するためである．二部グラフはデータセットとメタ情報間の多対多関係を表現したものであるが，データセット同士の距離は明示的に表現されていない．そこでランダムウォークを用いてデータセット同士の関係を系列情報として抽出し，埋め込みを行うことで距離を計算可能にする．

本手法のランダムウォークでは，D–T および D–W はいずれも「データセット→メタ情報→データセット→...（$d_0 \rightarrow t_1 \rightarrow d_1 \rightarrow t_2 \rightarrow d_2 \rightarrow \ldots$）」の交互遷移のみを許容するため，型制約付きランダムウォー

ク (Type-Constrained Random Walk) を採用する．これにより，直接の共通タグや共通語を持たない場合でも，中間ノードを介して意味的に近いデータセット同士が系列内で近接するようになる．これは DeepWalk や node2vec が一般のグラフで実現している高次近傍の共起抽出を二部グラフに適用したものであり，メタデータ由来のデータセットの意味構造を統計的に取り出すための処理である．

タグビューでは，TF–IDF および PPMI による重み付けの後，行方向に正規化を行うことで遷移確率行列 $P_{\mathrm{D}\to\mathrm{T}}$ および $P_{\mathrm{T}\to\mathrm{D}}$ を構成する．これにより，頻出タグの過度な影響を抑えつつ，タグの情報量を反映した確率的遷移が得られる．ランダムウォークはこれらの遷移行列を交互に適用し，データセットからタグへ，タグからデータセットへと遷移することで文を生成する．最終的にはデータセットノードのみを抽出し，共通するタグを有する確率の高いデータセット同士が近くに現れやすいコーパスを形成する．

テキストビューのランダムウォークもタグと同様であるが，中間ノードとして語を利用する点が異なる．語彙はタグと比較してはるかに多様であるため，BM25 重みに基づく遷移確率を設計することで，文書固有の特徴語を適切に重み付けした類似性を抽出する．語彙の共有はタグの共有よりも粒度の細かいデータセット同士の意味的関連性を有しているため，テキストビュー由来の文は，データセット間の潜在的トピックや内容構造をより詳細に表現できる．

こうして生成されたシーケンスは，タグビューとテキストビューで異なる性質を持つ．タグビューは主にデータセットの領域横断的，カテゴリー的な類似性を，テキストビューは文脈的，データの内容に踏み込んだ類似性をそれぞれ捉えることができる．両者は互いに補完的であり，両方を SGNS に入力することで，タグでは粗すぎる情報，テキストでは冗長になりがちな情報の両方をバランスよく抽出できることが期待できる．

3.4 SGNS によるビュー別データセット埋め込み

本節では，前節で生成したシーケンスを用いて，タグビューおよびテキストビューに対して SGNS を適用し，各データセットの埋め込み表現を学習する方法を述べる．SGNS を用いる理由は，シーケンスに潜む高次の共起構造を低次元の連続表現に圧縮できる点にある．SGNS は中心ノードとその周囲の文脈ノードの共起を最大化しつつ，負例サンプリングによりノイズとなる関係の出現を抑制するため，二部グラフ由来の複雑な構造パターンを柔軟に学習できる．

まず，ランダムウォークによって得られたデータセット列 $[d_0, \ldots, d_{L-1}]$ に対し，スライディングウィンドウを用いて，各中心ノードの前後 w ステップ以内に現れるノードを文脈ノードとすることで正例ペア $(\mathrm{center}, \mathrm{context})$ として取り出す．一方，ウィンドウの中に現れないノードを，シーケンス中の出現頻度に基づいて構成した分布からサンプリングし，負例ノードととして取り出す．高頻度ノードに過度に偏らないよう，頻度の 3/4 乗に比例する平滑化分布を用いる．これにより，本来共起しにくいノードペアを明示的に負例として与えることができる．

続いて，各データセットノートに対して「入力ベクトル」と「出力ベクトル」の 2 種類の埋め込みを保持し，バッチ単位で中心・文脈・負例ノードの ID をまとめて取得する．中心ベクトルと文脈ベクトルの内積が大きくなるように，また中心ベクトルと負例ベクトルの内積が小さくなるように損失を計算し，埋め込み行列を同時に更新する．実装においてはベクトル化した `logsigmoid` により一括計算し，タグビューから得たシーケンスとテキストビューから得たシーケンスに対してそれぞれ独立に SGNS を学習することで，タグ共起に基づく埋め込み Z_{tag} と，説明文共起に基づく埋め込み Z_{text} を得る．どちらも同じデータセット集合を行方向に共有するが，表現している意味的側面はビューごとに異なる．

実験で用いる Meta Kaggle データでは，ノード数（データセット数）が数十万，ランダムウォークで生成したシーケンスが数百万規模となるため，学習には膨大な計算量を要する．本研究では，PyTorch DistributedDataParallel を用いたデータ並列学習および混合精度訓練を併用し，ウォーク列を分割して各プロセスに割り当てることで，大規模コーパスに対しても効率的な学習を実現した．学習後は入力埋め込み行列 E_{in} をデータセットの最終表現として採用し，正規化したうえで近似最近傍探索に入力することで，ビューごとのデータセット類似グラフを構築する．

3.5 行動ビュー（Behavior View）の構築

行動ビューは，データセットそのものの内容やデータセットの内容について記述されたメタデータではなく，利用者の行動パターンや作成者の属性を通じて得られる関係性を捉えるビューである．具体的には，(1) データセットの作成者，組織 ID に基づく共属関係，および (2) 閲覧数，ダウンロード数，投票数，利用回数，経過日数などの利用ログに基づく行動類似度を扱う．これらは，実データ内容や一般的なメタデータとは異なる「誰が，どのような目的で利用しているか」というデータセットの機能的，文脈的な側面を反映しており，タグやテキストでは捉えられない重要な情報源となる．

行動ビューの構築には，まず共属関係グラフと利用行動グラフを別々に構成する．共属関係グラフでは，同一ユーザ（または組織）が作成したデータセット同士を強く連結するように重み付けする．一方，利用行動グラフは，閲覧，ダウンロード，投票，利用といった行動系列に基づく共通パターンを反映し，データセット間の利用パターンの類似性を表す．

続いて，行ごとの特徴強度（ノルム・分散など）を指標とし，各行に対して「どのシグナルをどれだけ信頼するか」を示す重みを算出する．すなわちデータセットごとに，共属情報と利用行動情報の寄与度を判断し，重み付き加重和を取ることで統合した行動ビューを作成する．その後，冗長なエッジを抑制するための行内 top-K の取得と行正規化を行い，最終的に行動ビューのデータセット類似グラフを得る．

3.6 3 つのビューの融合

最後に，タグビュー，テキストビュー，行動ビューの 3 つの類似グラフを統合するマルチビュー融合を行う．提案手法の特徴は，各データセットごとに 3 つのビューの信頼度を推定し，ビュー間で重みを動的に調整する適応的融合（Adaptive Fusion）を採用している点にある．このとき，まず Fused3-RA 手法で 3 つのビューのみから構築した類似度行列と，Fused3-RRF 手法にてタグ・説明文・作成者に基づいて候補近傍を再スコアリングした類似度行列を得る．最後に，両者の長所を合わせるために，Fused3-Blend 手法（$S_{\text{blend}} = (1-\eta)S_{\text{RA}} + \eta S_{\text{RR}}$ によって計算）を用い，ビュー特性の差異やスケールの不一致を吸収しつつ統合類似度行列を得る．

この融合行列は，各データセットに対して「タグが示す主題情報」,「テキストが表す意味情報」,「行動が示す利用文脈・機能特性の情報」の各ビューを総合的に結びつけたものであり，単一ビューでは捉えきれないデータセットに関するより広い観点に基づいたデータセット同士の類似性評価を可能にする．特に，タグが欠落したデータセットや，説明文が短いデータセットについては，行動ビューが代替情報として機能する．また，一方で行動情報の偏りはタグとテキストビューによって補正される．このような相補的なビューの融合により，ノイズの影響を抑えつつ，強いビューが弱いビューを支える構造を実現できる．

4 実験設定

4.1 データセットと設定

本実験では，メタデータのみを用いた類似度推定を検証するという目的のため，Meta Kaggle データセットを採用した[1]．これは Kaggle が公開しているメタデータの集合であり，データセットの識別子，タグ，タイトルとサブタイトル，概要説明，作成日時に加え，閲覧数，ダウンロード数，投票数，Kernel 利用回数といった利用統計情報を含む．全体で約 52 万件のデータセットのメタデータと，597 種類の異なるタグが存在する．

ランダムウォークから得た各データセット列については，長さ 40 のシーケンスがタグビューで約 200 万件，テキストビューで約 400 万件となった．また，実験では，評価指標との整合性とマルチビュー構造の保持とのバランスが最も良かったことから，代表値として $\eta = 0.3$ を採用した．

4.2 実験手順

元データには「データセット間の真の類似度」（Gold Label）を表す正解データが存在しないため，代替指標として以下の (1) タグ類似度，(2) テキスト類似度，(3) 行動類似度を用いることでデータの類似度を計算する提案手法の評価を行う．

1. **タグ類似度**: 前処理・正規化済みのタグ集合を用いる．2 つのデータセットが共有するタグに対し，IDF に基づく重みを付与して加算し，その値をタグ類似度とする．これにより，頻度が低いが識別力の高い概念が重み付けされ，データセットの主題的な近さを捉える．

2. **テキスト類似度**：Title, Subtitle，Description を連結し，BM25 ベクトルに変換する．2 つの BM25 ベクトルのコサイン類似度を計算し，必要に応じて閾値で二値化して MAP などの指標に用いる．データセットの意味的な類似度を捉える．

3. **クリエイター類似度**：CreatorUserId を用い，同一ユーザであれば 1，それ以外は 0 とする単純な二値の関連度とする（ただし，欠損は –1）．作成者が同一であればデータセットの対象や設計が類似しやすいという点を踏まえた関連指標である．

比較実験では，単一ビュー（タグ PPMI ＋コサイン，テキスト BM25 ＋コサイン類似度，行動特徴のコサイン類似度），単純なランキング融合手法（RRF, CombSUM），そして本研究のマルチビュー融合手法（Fused3-RA, Fused3-RRF, Fused3-Blend）を評価対象とする．最終的な総合スコアは，タグ類似度，テキスト類似度，クリエイター類似度にそれぞれ 0.6 / 0.3 / 0.1 の重みを付与した統合指標とし，検索性能を比較する．

さらにアブレーション解析により，各ビューの寄与度を分析する．具体的には，タグ＋テキストのみで行動

[1] 用いたデータセットの最終更新日は 2025 年 10 月 25 日である．

表 1: Result of Comparison with Baseline

Method	nDCG@20	MAP@20	MRR@20	P@20	R@20
Tag-SGNS	0.0327	0.0883	0.0982	0.0328	0.0001
Text-SGNS	0.0329	0.0879	0.0978	0.0326	0.0000
Tag-BM25-Cos	0.1403	0.1854	0.1948	0.1364	0.0169
Text-PPMI-Cos	0.7721	0.7700	0.7681	**0.7725**	0.0232
Behavior-Eng-Cosine	0.0516	0.1000	0.1186	0.0499	0.0042
Fusion-RRF	0.3799	0.4573	0.6453	0.3694	0.0375
Fusion-CombSUM	0.2490	0.3829	0.4108	0.2359	0.0439
Fused3-RA	0.8746	0.8717	0.8607	0.3217	0.8376
Fused3-Blend-eta0.30(Ours)	**0.9119**	**0.9128**	**0.8960**	0.5898	**0.8429**

ビューを使わないなど，あるビューを除外することで，提案手法であるマルチビュー融合の効果を確認する．

4.3 評価指標の設計

実験では，次の 5 つの評価指標を用いて各手法の性能を比較する．

1. **nDCG@20**：各サンプルについて，上位 20 件の近傍において，タグ共有・テキスト BM25 コサイン類似度が高い・クリエイターが同一などの関連候補がどれだけ上位に並んでいるかを，位置に応じた重みで評価し，正規化した指標である．値が大きいほど，関連データセットが優先的に提示されていることを意味する．
2. **MAP@20**：各サンプルの上位 20 件リスト内で，関連項目が出現した時点の精度を計算し，その平均を全サンプルで平均したもの．値が高いほど，関連データセットがリストの上位に密に出現していることを示す．
3. **MRR@20**：各サンプルについて，上位 20 件内で最初に現れる関連近傍の順位の逆数を取り，それを平均した指標である．値が大きいほど，ユーザが最初に目にする候補が関連データセットである可能性が高いことを表す．
4. **Precision@20**：上位 20 件中に含まれる関連項目の割合．
5. **Recall@20**：上位 20 件に含まれる関連項目数が，全関連項目集合のうちどの程度を占めるかを表す．値が大きいほど，関連データセットを取りこぼさずに網羅できていることを示す．

5 実験結果と考察

5.1 ベースライ方法との比較実験

表 1 は，全手法の統合指標に基づく順位を示しており，本研究の手法 Fused3-Blend-eta0.30 が他の全てを大きく上回っていることを表している．この順位から，いくつかの階層的な傾向を読み取ることができる．第一に，融合手法はほとんどの指標において，総じて単一ビュー手法より優れた性能を有している．第二に，適応的融合（Fused3-RA および改良版 Fused3-Blend）は，単純なランキング融合（RRF, CombSUM）より高い性能を発揮している．第三に，行動ベースの単一ビューとテキストベースの単一ビューがいずれも上位に位置しており，データセット推薦においてはタグだけでなくテキスト情報も同程度に重要であることが分かる．

5.2 アブレーション実験

アブレーション実験では，融合手法からビューを個別に除去し，対応する単一ビューのベースラインと比較することで，異なるビューを融合することの有効性を検証する．

5.2.1 タグビューの検証

表 2 より，タグビューのみを用いる Tag-SGNS の nDCG@20 は 0.0300 と低く，タグ共有だけではデータセット類似性を十分に表現できないことが分かる．これは，SGNS が共起統計に依存する一方で，異なるタグ数が 597 個と少なく，分布も偏っており，多くのタグペアで共起回数が極めて低いためである．その結果，意味的な関連とノイズの切り分けが難しく，得られた埋め込みベクトルの類似度表現能力が弱くなってしまっていると考えられる．

表 2: Result of View-Ablation Experiments

Method	nDCG@20	MAP@20	MRR@20	P@20	R@20
Tag-SGNS	0.0300	0.0804	0.0894	0.0299	0.0000
Text-SGNS	0.0003	0.0011	0.0011	0.0003	0.0000
Text-BM25-Cos	0.1425	0.1675	0.1514	0.1400	0.0409
Behavior-Eng-Cosine	0.0081	0.0143	0.0146	0.0076	0.0017
(Behavior-Similarity-only)	0.8654	0.8780	0.8775	0.3420	0.7953
Fusion-RRF (Tag + Text)	0.0277	0.0435	0.0491	0.0205	0.0161
Fused3-RA (Tag + Text)	0.8330	0.8457	0.8370	0.3132	0.7890
Fused3-Blend (Tag + Text)	**0.8944**	**0.8503**	**0.8395**	**0.3199**	**0.8000**
Fusion-RRF (Tag + Behavior)	0.1514	0.2392	0.2724	0.1368	0.0356
Fused3-RA (Tag + Behavior)	0.4173	0.5855	**0.5998**	0.2290	0.3035
Fused3-Blend (Tag + Behavior)	**0.7197**	**0.6323**	0.4331	**0.3298**	**0.3709**
Fusion-RRF (Text + Behavior)	**0.3591**	**0.4375**	**0.6114**	**0.3482**	**0.0023**
Fused3-RA (Text + Behavior)	0.1097	0.2358	0.2667	0.0951	0.0002
Fused3-Blend (Text + Behavior)	0.0712	0.1874	0.1916	0.1278	0.0004

一方，マルチビュー融合手法はベースラインを大きく上回る．タグビューとテキストビューを融合した Fusion-Blend は nDCG@20 を 0.8330 まで向上させ，Tag-SGNS 比で約 30 倍となり，他の指標でも大幅な改善が見られる．タグビューと行動ビューを組み合わせた場合も同様に性能が向上している．これは，タグビューが明示的なデータセットの主題を，テキストビューが潜在的な意味を，行動ビューがユーザー行動パターンをそれぞれ捉え，これらを融合することでデータセット間の類似性をより包括的かつ精度高く表現できることを示している．

5.2.2 テキストビューの検証

テキストビューでも顕著な差が見られる（表 2）．テキストビューのみを用いた Text-SGNS の nDCG@20 はほぼ 0 であり，データセット類似性をほとんど捉えられていない．これは，説明テキストが長いうえにノイズも多く，D–W の二部グラフ上の共起構造が疎であったため，SGNS で有効な統計シグナルが得にくかったことが原因と考えられる．テキストの内容を BM25 ベクトル化しコサイン類似度で評価しても，改善には限界があり，単一ビューでは 0.14 程度にとどまった．

一方，テキストビューとタグ・行動ビューを組み合わせた融合手法は大きな性能向上を示した．タグビュー単体のシグナルは弱くても，他ビューと融合することで，タグを持たないデータセットに対してテキストビューと行動ビューが代替的な関連性シグナルを提供し，タグの不足を補えていることが分かる．

さらに，テキストビューとタグビューの融合は大きな性能の向上を示したのに対し，テキストビューと行動ビューのみの融合は相対的に向上の度合いが小さく，データセット類似性の表現にはタグビューの寄与が不可欠であることが確認された．この結果は，メタデータを異なるビューに分離して扱う本稿の設計が妥当であることを裏づけている．

5.2.3 行動ビューの検証

はじめに行動ビューのみに基づいてデータセット類似度を計算すると，性能は相対的に低くなることが分かった．続いて，クリエイター類似度のみを用いてデータセット間の類似度を直接評価し直すと，高スコアが得られ，行動ビュー自体は「ユーザ活動の観点からの類似性」を的確に捉えていることが確認できた．ただし，行動ビューのみではデータセットが有する内容や主題といった側面は表現できないため，タグ・テキストビューと組み合わせることで情報を補完することが重要になる．

表 2 からも，行動ビューとタグ・テキストビューの融合が大きな性能向上に寄与していることが分かる．強いシグナルを持つビューが弱いビューを補完し，ビュー間の寄与を動的に調整することで全体性能が底上げされるという，マルチビュー融合の本質的な意味であり，単純な平均ではなく，ビュー間の寄与を調整することが各種指標を同時に押し上げているものと考えられる．

6 おわりに

本研究は，データセットの内容，すなわち実データに直接アクセスすることなくデータセット間の類似度を計算するために，タグ，説明テキスト，ユーザ行動

という 3 種類の異なるメタデータビューによるマルチビュー表現と適応的融合の新しいフレームワークを提案した．このようなメタデータベースのマルチビュー類似度学習は，実際のデータ市場やオープンデータプラットフォームにおけるデータ発見と再利用支援において，いくつかのメリットがある．第一に，データ内容にアクセスせずに類似データセット候補を提示できるため，プライバシーやコンプライアンス上の制約が厳しい環境でも，利用者が候補を絞り込むための検索エンジンとして機能し得る．第二に，異なる性質のメタデータを統合することで，単一のキーワード検索や単一ビューでは見落とされがちな関連データセットや代替データセットを推薦でき，様々な下流タスクを支援できる．第三に，提案手法で得られた類似度グラフは，クラスタリングやトピック分析，データカタログの自動整理などにも利用可能であり，データエコシステムの基盤となることが期待できる．

一方で，本研究にはいくつかの限界も存在する．まず，実験は Kaggle という単一のデータプラットフォームのメタデータに基づいており，ドメインや言語，メタデータスキーマの異なるデータカタログに対しても同様の性能が得られるかどうか検証が必要である．また，評価にはタグ・テキスト・作成者情報から構成した代替指標を用いており，真の利用者満足度やタスク達成度を直接測定しているわけではない．そのため，真の類似度を得るためには，表形式データ向け言語モデル（TaLMs）などを援用した類似度比較も必要となる．さらに，ユーザ行動ビューは多数のデータセットを提供する作成者や人気のデータセットにバイアスを受けやすく，公平性・多様性の観点からの検討も重要である．今後は，より大規模かつ多様な外部データポータルを対象とした実験を行うとともに，オンライン評価やユーザスタディによる実運用下での有効性検証，行動ログのバイアスを緩和する融合戦略や，新たなメタデータ種別を取り込んだ拡張モデルの検討も重要な課題である．

謝辞

本研究は JSPS 科研費（JP25K00153）の助成を受けました．

参考文献

[1] Azcoitia, S. A., Laoutaris, N.: A Survey of Data Marketplaces and Their Business Models, *SIGMOD Record*, Vol. 51, No. 3, pp. 18–29 (2022)

[2] Chen, Z.: Challenges and Progress in Dataset Search, *Proc. 8th Symposium on Future Directions in Information Access*, (2020)

[3] Nargesian, F., et al.: Dataset Discovery in Data Lakes (D3L), *arXiv preprint*, (2020)

[4] Lv, Q., Josephson, W., Wang, Z., Charikar, M., Li, K.: Ferret: A Toolkit for Content-Based Similarity Search of Feature-Rich Data, *Proc. EuroSys 2006*, pp. 317–330 (2006)

[5] Yan, Z., Tang, D., Duan, N., Bao, J., Lv, Y., Zhou, M., Li, Z.: Content-Based Table Retrieval for Web Queries, *Neurocomputing*, Vol. 349, pp. 183–189 (2019)

[6] 作本 猛, 早矢仕 晃章, 坂地 泰紀, 野中 尋史: 類似データセット発見課題における詳細なデータセット分類に基づいた有効性の評価, 言語処理学会 第 *29* 回年次大会発表論文集 (2023)

[7] Ravishankar, T. N., Shriram, R.: Metadata Based Clustering Model for Data Mining, *Journal of Theoretical and Applied Information Technology*, Vol. 67, No. 1, pp. 59–67 (2014)

[8] Bernhauer, D., Nečaský, M., Škoda, P., Klímek, J., Skopal, T.: Open Dataset Discovery Using Context-Enhanced Similarity Search, *Knowledge and Information Systems*, Vol. 64, No. 12, pp. 3265–3291 (2022)

[9] Sakumoto, T., Hayashi, T., Sakaji, H., Nonaka, H.: Metadata-based Clustering and Selection of Metadata Items for Similar Dataset Discovery and Data Combination Tasks, *IEEE Access*, Vol. 12, pp. 40213–40224 (2024)

[10] Kumar, A., Rai, P., Daumé, H. III: Co-regularized Multi-view Spectral Clustering, *Advances in Neural Information Processing Systems 24 (NeurIPS 2011)*, pp. 1413–1421 (2011)

[11] Wang, B., San Lucas, A. G., Shah, N., Man-Child, E., Kantarcioglu, M., et al.: Similarity Network Fusion for Aggregating Data Types on a Genomic Scale, *Nature Methods*, Vol. 11, No. 3, pp. 333–337 (2014)

[12] Gönen, M., Margolin, A. A.: Localized Data Fusion for Kernel k-Means Clustering with Application to Cancer Biology, *Advances in Neural Information Processing Systems 27 (NeurIPS 2014)*, pp. 1305–1313 (2014)

高等学校情報科「情報I」における各教科書の特徴の可視化

Visualization of Characteristics in High School Information Science Textbooks for “Information I”

菊谷 和也[1]　　*笹嶋 宗彦[1]
Kazuya Kikutani[1]　　Munehiko Sasajima[1]

[1] 兵庫県立大学
[1] University of Hyogo

Abstract: Starting in the 2025 academic year,“Information”was newly added to the Common Test for University Admissions, but the level of questions and the key areas of learning content have not been clearly defined. This study analyzes the commonalities and differences in content levels across different high school Information Science“Information I”textbooks. By comparing the vocabulary and frequency lists for“Information I”textbooks published by the Information Processing Society of Japan with the key terms in each textbook, characteristics of each textbook became apparent, such as one textbook having a higher level of programming education compared to others.

1　はじめに

高等学校では2018年3月に新しい学習指導要領である平成30年告示高等学校学習指導要領[1]（以下,「新学習指導要領」と表記）が告示され，2022年4月から年次進行で実施が始まっている．この新学習指導要領では，これまで選択必修科目であった「社会と情報」と「情報の科学」の2科目が統合され，新科目「情報I」が共通必履修科目として設けられた.「情報I」で学ぶ内容には，プログラミング，モデル化とシミュレーション，ネットワークやデータベースの基礎など，これまでの高等学校教育では扱われてこなかった分野も含まれている．また,「情報I」は2025年に実施された大学入試共通テストにおいてプログラミングを含む「情報」として出題された．しかし,「令和9年度大学入学者選抜に係る大学入学共通テスト出題教科・科目の出題方法等」[2]において,「情報」の出題範囲が明記されておらず，また高等学校における学習内容の重点化の基準も明確に定まっていない[3].

「情報I」の教科書は複数出版されており，それぞれ新学習指導要領に沿って内容が改定されているが，教科書間で掲載されている語句には違いがある．実際に12種類の教科書の索引に掲載されている語句を調査した中園の研究[4]では，索引語句数が189語から612語までの開きがあり，教科書によって大きく異なることを示している．また，2つの異なる出版社の教科書に含まれている重要語を調査した著者らの研究[5]でも，重要語数が140語から426語と3倍近くの開きがあることが確認できている．なお，重要語の定義については，第2節で述べる．このことから，各教科書の内容構成には一定のばらつきがあり，教科書ごとの特徴を体系的に整理することが必要であるといえる．

そこで，本研究では高等学校情報科「情報I」の教科書を対象として，各教科書における用語の分布や特徴を可視化し，教科書間の構成上の差異を分析することを目的とする．特に，情報処理学会が公開している「情報科全教科書用語リスト」[6]を参照し，教科書内の語句を分野別に照合することで，学習指導要領の4大領域に対する各教科書の重点傾向を明らかにする．

本稿では，教科書に記載された概念と「情報科全教科書用語リスト」との照合結果をもとに，教科書間での用語の扱われ方や分布の違いについて報告する．第2節では，教科書の特徴を表す用語の抽出および分類の基準について述べ，第3節では教科書の特徴を可視化する方法を示す．第4節では，可視化を行った5冊の教科書を対象とした分析結果を報告し，第5節で本稿のまとめと今後の展望について述べる．

2　教科書の特徴を表す用語の基準

本研究では，教科書に記載された重要な概念を適切に抽出し，それらを比較可能な形で整理することが重要であると考える．本節では，教科書の特徴を分析する上で基礎となる用語の基準について定義する．

*連絡先：兵庫県立大学情報科学研究科
〒651-2197 兵庫県神戸市西区学園西町8丁目2-1
E-mail:ad24b015@guh.u-hyogo.ac.jp

2.1 重要索引語

東京都教育委員会が公開している「令和 4 年度使用都立高等学校および都立中等教育学校（後期課程）用教科書教科別採択結果（教科書別学校数）」[7] のデータによると，都立高等学校において採択数が最も多い教科書は「最新情報 I」（実教出版，2022）[8] である．この教科書を基準として，用語抽出の定義および分析方法を検討した．

本研究では，教科書内で太字表示されている用語を「重要語」とし，教科書において特に重要とされる用語として定義している．「最新情報 I 」には 426 語の重要語が確認されているが，索引に記載されている用語数は 678 語であり，重要語と索引語句の間に違いがある．この違いについて調査したところ，重要語には教科書の重要な概念とは必ずしも関連しない用語も含まれていることが明らかになった [9]．

以上の結果を踏まえ，本研究では，重要語のうち索引にも掲載されている用語を「重要索引語」と定義する．重要索引語は，執筆者が学習上重要とみなしていると考えられる語であり，教科書がどの内容を重視しているかを示す指標として有効であると考えられる．本研究では，この重要索引語を用いて，各教科書間における語彙の分布や重複の傾向を比較し，教科書ごとに重点を置いている内容の違いを明らかにするとともに，教科書の特徴を示す指標として活用する．

2.2 教科書間における重要索引語の分布

教科書に掲載されている用語数および重要索引語の重複度を比較するために,「最新情報 I」（実教出版，2022）[8],「高校情報 I Python」（実教出版，2022）[10],「高等学校情報 I」（数研出版，2022）[11],「新編情報 I」（東京書籍，2022）[12],「情報 I」（日本文教出版，2022）[13] の 5 冊を対象とした分析を行った．

図 1 は，各教科書における索引語数，重要語数，および重要索引語数を示している．索引語は 302 語から 685 語，重要語は 139 語から 437 語，重要索引語は 133 語から 387 語と幅があり，教科書間で大きな差が確認された．また，どの教科書においても「索引語＞重要語＞重要索引語」の順に語数が減少しており，索引に掲載されたすべての語が太字で強調されているわけではないことが分かる．これは，執筆者が強調語として扱う語の基準が異なっている可能性を示している．特に,「高校情報 I Python」と「最新情報 I」はいずれも実教出版から刊行されているが，用語数に若干の違いが見られることから，同一出版社内でも編集方針や指導観点の違いが存在する可能性がある．

次に，重要索引語がどの程度教科書に共通して掲載されているかを分析した結果を図 2 に示す．5 冊すべての教科書に共通して掲載されていた重要索引語は 39

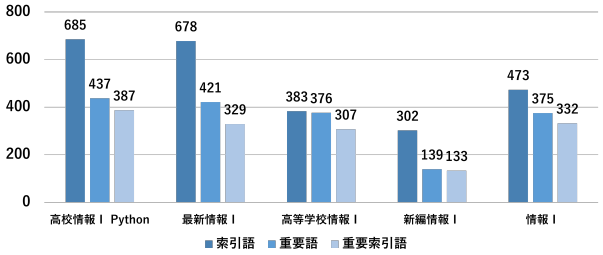


図 1: 教科書ごとの用語数

語（4.43 %）であり，非常に少数であることが分かった．共通して確認された用語には,「プログラム」,「シミュレーション」,「ソーシャルエンジニアリング」など，情報 I の学習内容の中心的概念が含まれている．一方で，1 冊の教科書のみに掲載されていた重要索引語は 556 語（63.25 %）と最も多く，各教科書が独自に重視している内容が多数存在していることが示唆された．

さらに，重要索引語の分布を見ると，2 冊または 3 冊で共通している用語も一定数存在し，部分的な重なりが見られる．これは，プログラミング，情報社会，ネットワークなどの特定の分野では出版社間で共通理解がある一方で，詳細な概念や強調の仕方には違いがあると考えられる．

以上の結果から，重要索引語は教科書間の内容の重点化や編集方針の差異を示す有効な指標であることが分かった．共通語の少なさは，教科書間で扱う概念の選定や重点づけが多様であることを示しており，学習内容の標準化という観点から今後の課題であると考えられる．

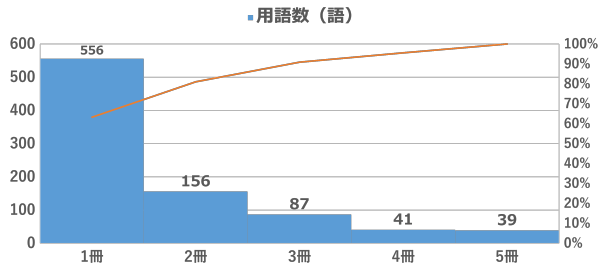


図 2: 掲載教科書数ごとの重要索引語の割合

3 教科書からの用語抽出法

本節では，教科書の構成や内容の特徴を可視化するために行った用語抽出と階層構造化の手法について述べる．本研究では，教科書の目次構造と本文を利用し，目次単位で出現する用語を対応付けることで，教科書内の概念構成を明確化することを目的とした．

まず，教科書から目次情報を抽出し，章・節・項・小項の 4 層構造をもつ目次ツリーを作成した．次に，教

科書本文を対象に，情報処理学会が公開している「情報科全教科書用語リスト」に掲載されている語を照合し，一致した語句のみを抽出した．このとき，本文中で太字で示されている用語は，執筆者が学習上特に強調している重要な語として扱い，重要索引語に分類した．また，抽出結果内ではこれらの語を「★」を付けて示した．

さらに，抽出された用語を「情報科全教科書用語リスト」に基づいて以下の 4 分類に整理した．

- **全教科書掲載**：全ての教科書に掲載されている
- **掲載数上位**：全教科書の 70 %以上
- **掲載数中位**：全教科書の 30 %より大きく 70 %より小さい
- **掲載数下位**：全教科書の 30 %以下

このように分類することで，教科書内の各目次項目における用語の出現傾向を比較し，共通的な学習内容が多い領域と各教科書固有の内容が多い領域を区別できる．この分類は，教科書間での内容の重点化や独自性を示す指標として用いる．

図 3 は,「最新情報 I」（実教出版，2022）の第 6 章「アルゴリズムとプログラミング」における抽出結果の一部を示している．黒色の項目は全教科書で共通して掲載されている用語，赤色は掲載教科書数上位（70 %以上），黄色は中位（30〜70 %未満），青色は下位（30 %以下）の用語を表す．また，★印は本文中で太字として強調されていた重要索引語である．

この図から,「アルゴリズム」「プログラム」「フローチャート」などの用語は多くの教科書に共通して登場し，情報 I の中核概念として扱われていることが分かる．一方で,「状態遷移図」などの用語は，一部の教科書にしか登場しておらず，特定の出版社が重視する内容であることが確認できる．これにより，同じ「アルゴリズムとプログラミング」という単元内でも，教材によって強調される表現手法や具体例が異なることが視覚的に把握できる．

さらに，階層構造に基づく可視化を行うことで，各用語がどの章節構造に位置づけられているかを明示でき，教科書内での概念の導入順序や扱い方を整理することができる．この手法を用いて，次節では 5 冊の教科書を対象として，抽出された用語の分布や重複の傾向を比較し，教科書間における内容構成の特徴を明らかにする．

4 用語抽出結果

前述した手法を用いて,「最新情報 I」（実教出版，2022）[8],「高校情報 I Python」（実教出版，2022）[10],「高等学校情報 I」（数研出版，2022）[11],「新編情報 I」

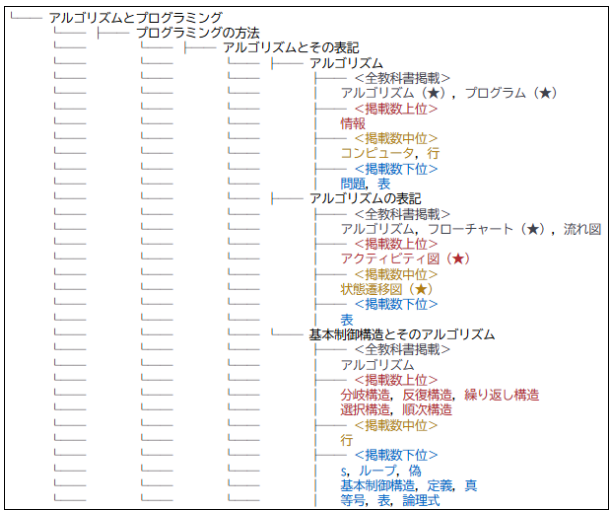


図 3: 「最新情報 I」における階層構造と掲載状況の可視化（一部抜粋）

（東京書籍，2022）[12],「情報 I」（日本文教出版，2022）[13] の 5 冊を対象とした可視化を行った．

4.1 高校情報 I Python

「高校情報 I Python」（実教出版，2022）は,「情報社会」,「情報デザイン」,「デジタル」,「ネットワーク」,「問題解決」,「プログラミング」の 6 章で構成されている．

表 1 は,「高校情報 I Python」における領域ごとの掲載教科書数および割合を示している．全体として,「コンピュータとプログラミング」および「情報通信ネットワークとデータの活用」の領域に属する用語数が多く，特に掲載数中位，掲載数上位の語が両領域に集中している．このことから，本教科書はプログラミングやデータ処理など，情報技術の実践的活用を重視した構成であることが分かる．一方で,「情報社会の問題解決」や「コミュニケーションと情報デザイン」領域の掲載数はやや低く，社会的課題の考察や情報発信の側面よりも，情報処理の技能的内容に重点が置かれていることが示唆される．

表 1: 「高校情報 I Python」における領域ごとの掲載教科書数及び割合

	情報社会の問題解決		コミュニケーションと情報デザイン		コンピュータとプログラミング		情報通信ネットワークとデータの活用	
	個数	%	個数	%	個数	%	個数	%
全教科書掲載	13	14.1	16	13.0	11	8.6	14	8.7
掲載数上位	30	32.6	26	21.1	32	25.0	48	29.8
掲載数中位	26	28.3	54	43.9	65	50.8	72	44.7
掲載数下位	23	25.0	27	22.0	20	15.6	27	16.8
合計	92	100	123	100	128	100	161	100

「コンピュータとプログラミング」に対応する章である「プログラミング」の分析結果からは，アルゴリ

ズム・基本構造・変数といった基礎概念から，配列・関数・探索／整列・設計手法，さらに可視化や計測制御といった応用内容へと，技能と抽象化を段階的に積み上げる体系的構成であることが確認できた．

特に図 4 に示す「オブジェクト指向プログラミング」では，教科書タイトルにもあるように「Python」を用いた実装を前提としており，他の教科書では触れられにくいクラス定義やメソッド，プロパティなどの概念を具体的に扱うことで，プログラミングの応用的理解を促す構成となっている．

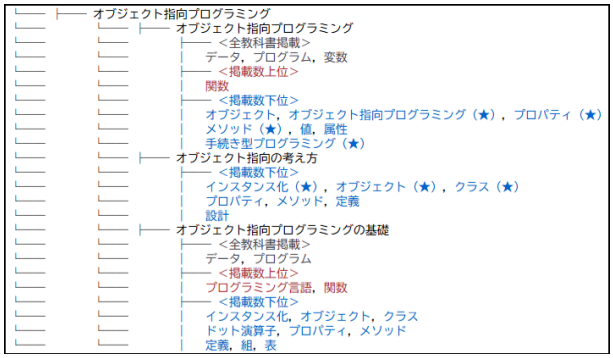


図 4: 「高校情報 I Python」における階層構造と掲載状況の可視化（一部抜粋）

4.2 最新情報 I

「最新情報 I」（実教出版, 2022）は,「情報社会と私たち」,「メディアとデザイン」,「システムとデジタル化」,「ネットワークとセキュリティ」,「問題解決とその方法」,「アルゴリズムとプログラミング」の 6 章で構成されている．

表 2 は，最新情報 I における領域ごとの掲載教科書数とその割合を示している．全体的に見ると,「情報社会の問題解決」および「コミュニケーションと情報デザイン」の領域に多くの用語が含まれている一方で,「情報通信ネットワークとデータの活用」の領域では掲載割合がやや低い傾向にある．特に掲載数中位，掲載数下位の語がこの 2 領域に集中していることから，これらの領域に属する内容は教科書全体で扱い方に差が生じていることが示唆される．

表 2: 「最新情報 I」における領域ごとの掲載教科書数及び割合

	情報社会の問題解決		コミュニケーションと情報デザイン		コンピュータとプログラミング		情報通信ネットワークとデータの活用	
	個数	%	個数	%	個数	%	個数	%
全教科書掲載	14	12.8	16	12.6	11	9.6	11	7.1
掲載数上位	36	33.0	28	22.0	30	26.1	41	26.6
掲載数中位	31	28.4	59	46.5	50	43.5	67	43.5
掲載数下位	28	25.7	24	18.9	24	20.9	35	22.7
合計	109	100	127	100	115	100	154	100

「情報通信ネットワークとデータの活用」に対応する目次である「問題解決とその方法」に関する分析結果からは，教科書内で「問題解決」の流れを段階的に示しつつ,「問題の明確化」「情報の収集」「解決案の決定」「評価」などの関連する語彙が体系的に配置されていることが確認できた．

特に図 5 にある「データ分析の手法」に関しては,「表計算ソフト」「関数」「相関」「回帰分析」などデータ分析手法に関わる語が多く出現しており，情報活用の過程を実践的に理解させる構成となっている．

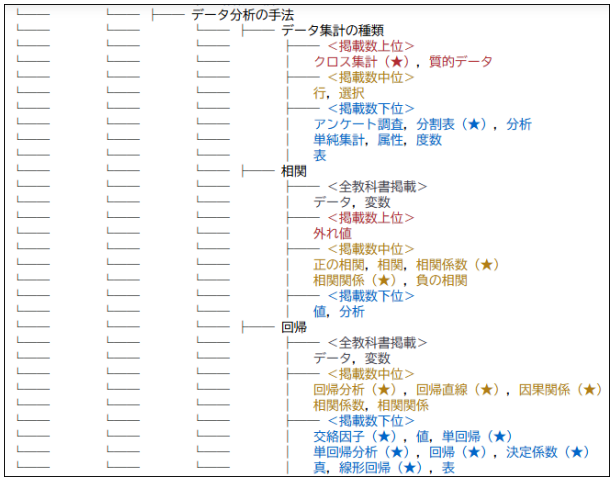


図 5: 「最新情報 I」における階層構造と掲載状況の可視化（一部抜粋）

4.3 高等学校情報 I

「高等学校情報 I」（数研出版，2022）は,「情報社会の問題解決」,「コミュニケーションと情報デザイン」,「コンピュータとプログラミング」,「情報通信ネットワークとデータの活用」の 4 章で構成されている．

表 3 は,「高等学校情報 I」における領域ごとの掲載教科書数およびその割合を示している．全体として,「情報通信ネットワークとデータの活用」と「情報社会の問題解決」の 2 領域で掲載数が多く，特に「情報通信ネットワークとデータの活用」では掲載数中位語が最も多くを占めており，データ活用やネットワーク関連の内容が多様な観点から扱われていることが分かる．一方で,「コンピュータとプログラミング」は掲載数全体に占める割合がやや低く抑えられており，プログラミングの基礎事項を重点的に取り上げる構成となっている．また,「コミュニケーションと情報デザイン」は掲載数中位語の割合が 37.5 %と高く，他の領域に比べて教科書間で扱い方にばらつきが見られることから，学習内容の多様性を意識した設計であるといえる．これらの傾向から，本教科書は社会的課題の解決や情報の共有・活用に重点を置き，実践的な情報活用能力の育

成を重視した構成であることが示唆される．

表 3: 「高等学校情報 I」における領域ごとの掲載教科書数及び割合

	情報社会の問題解決		コミュニケーションと情報デザイン		コンピュータとプログラミング		情報通信ネットワークとデータの活用	
	個数	%	個数	%	個数	%	個数	%
全教科書掲載	16	13.4	18	13.2	11	12.1	17	10.2
掲載数上位	38	31.9	34	25.0	19	20.9	45	27.1
掲載数中位	27	22.7	51	37.5	40	44.0	67	40.4
掲載数下位	38	31.9	33	24.3	21	23.1	37	22.3
合計	119	100	136	100	91	100	166	100

「情報通信ネットワークとデータの活用」に関する分析結果からは，ネットワークの基礎から始まり，アドレッシングと Web の仕組みを経て，通信の信頼性・暗号化による保護へと理解を段階的に深め，最終的にデータベースと各種データ分析へ接続する流れが確認できた．この構成は，通信路で生じるデータの生成・伝送・蓄積・活用までの一連のプロセスを俯瞰させる設計であると考えられる．

特に図 6 にある「データベース管理システムの機能」に関しては,「トランザクション」「ロールバック処理」「ユーザ認証」「実行権」「資源管理」など掲載数下位の語が集中的に現れており，一部の教科書が権限制御・運用管理といった実務寄りの概念まで踏み込んでいることが示唆される．

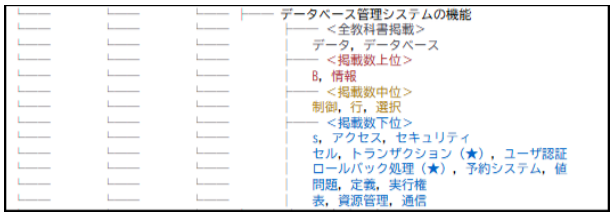


図 6: 「高等学校情報 I」における階層構造と掲載状況の可視化（一部抜粋）

4.4 新編情報 I

「新編情報 I」（東京書籍，2022）は,「情報で問題を解決する」,「情報を伝える」,「コンピュータを活用する」,「データを活用する」,「活動して提案する」の 5 章で構成されている．

表 4 は,「新編情報 I」における領域ごとの掲載教科書数および割合を示している．全体として,「情報社会の問題解決」と「情報通信ネットワークとデータの活用」の 2 領域で掲載数が多く，特に後者は掲載数上位語が最も多くを占めており，ネットワークやデータ活用に関する内容が充実していることが分かる．一方,「コンピュータとプログラミング」の割合は比較的低く，プログラミングの基礎事項を中心に抑えた構成となっている．また,「コミュニケーションと情報デザイン」は掲載数中位語の割合が 33.8 %と高く，他領域に比べて教科書間で扱い方にばらつきがみられる．これらの傾向から,「新編情報 I」は情報の利活用やネットワーク理解を重視しつつ，実践的な問題解決を通して情報社会での活用力を育成することを目的とした内容構成であると考えられる．

表 4: 「新編情報 I」における領域ごとの掲載教科書数及び割合

	情報社会の問題解決		コミュニケーションと情報デザイン		コンピュータとプログラミング		情報通信ネットワークとデータの活用	
	個数	%	個数	%	個数	%	個数	%
全教科書掲載	11	14.7	14	19.7	10	16.7	9	10.6
掲載数上位	27	36.0	20	28.2	19	31.7	33	38.8
掲載数中位	19	25.3	24	33.8	20	33.3	28	32.9
掲載数下位	18	24.0	13	18.3	11	18.3	15	17.6
合計	75	100	71	100	60	100	85	100

「コミュニケーションと情報デザイン」に対応する目次である「情報を伝える」に関する分析結果からは，コミュニケーションの歴史的変化からデジタル化の仕組み，情報デザインやユニバーサルデザインの概念までを段階的に学習し，情報の表現・伝達・共有のプロセスを体系的に理解させる構成となっていることが確認できた．また，数値・音・画像・動画といった多様な情報表現を通じて，情報の受け取り方や伝え方の多様性を意識させる展開が見られた．

特に図 7 にある「デザイン思考に沿った制作の流れ」に関しては,「コンテンツ」,「デザイン思考」,「ポスター」,「分析」,「問題」,「設計」など掲載数下位語が多く含まれており，実践的な制作活動や課題解決型の学習場面を意識した構成であることがうかがえる．このことから，本書では単なるデザイン技法の理解にとどまらず，問題発見から表現・発信までの一連の創造的プロセスを重視した指導設計がなされていると考えられる．

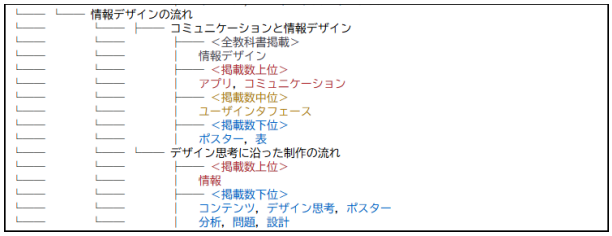


図 7: 「新編情報 I」における階層構造と掲載状況の可視化（一部抜粋）

4.5 情報 I

「情報 I」（日本文教出版，2022）は,「情報社会の問題解決」,「コミュニケーションと情報デザイン」,「コンピュータとプログラミング」,「情報通信ネットワークとデータの活用」の 4 章で構成されている．

表 5 は,「情報 I」における領域ごとの掲載教科書数およびその割合を示している．全体として,「情報通信ネットワークとデータの活用」の掲載語数が最も多く,他の領域に比べて内容の広がりが大きいことが分かる．この領域では掲載数中位，掲載数下位に分類される語の割合が高く，教科書ごとに扱い方や詳細度にばらつきがあることが示唆される．一方で,「コンピュータとプログラミング」は 117 語と全体の中では中程度の分量であるが，掲載数中位語が 43.6 %と高く，プログラミングの基礎事項を中心に統一的な内容が多く見られた．これらの傾向から，本教科書は「情報通信ネットワークとデータの活用」を中心に情報の循環的な利用を重視していることが確認できる．

表 5: 「情報 I」における領域ごとの掲載教科書数及び割合

	情報社会の問題解決		コミュニケーションと情報デザイン		コンピュータとプログラミング		情報通信ネットワークとデータの活用	
	個数	%	個数	%	個数	%	個数	%
全教科書掲載	13	13.5	17	15.7	12	10.3	11	6.5
掲載数上位	32	33.3	25	23.1	29	24.8	50	29.8
掲載数中位	25	26.0	35	32.4	51	43.6	59	35.1
掲載数下位	26	27.1	31	28.7	25	21.4	48	28.6
合計	96	100	108	100	117	100	168	100

「コンピュータとプログラミング」に関する分析結果からは，コンピュータの基本構成や CPU・メモリなどの動作原理を理解した上で，アルゴリズムの表現方法やプログラミングの構成要素，データ構造を段階的に学習し，最終的に Python による実装やシミュレーションへと展開する流れが確認できた．この構成は，計算機の仕組みからアルゴリズム設計，実装・応用へと知識と技能を統合的に習得させる体系となっている点に特徴がある．

特に図 8 にある「CPU による演算のしくみ」に関しては,「XOR」「論理ゲート」「真理値表」など掲載数下位語が多く見られ，論理回路や計算機構造の詳細を扱う高度な内容となっている．これらの語は他教科書では省略される傾向にあるため，本書ではプログラミング的思考を支えるハードウェア理解を重視し，論理演算とプログラム処理の関連を深く捉えさせようとする意図が読み取れる．

5　おわりに

本研究では，高等学校情報科「情報 I」の教科書を対象に，用語の分布と特徴を可視化し，教科書間における構成上の違いを分析した．この分析により，各領域で重視されている内容や，用語の扱われ方の傾向を俯瞰的に把握することができた．

今後は，情報処理学会が公開する「情報科全教科書用語リスト」を基準として，各教科書における用語群

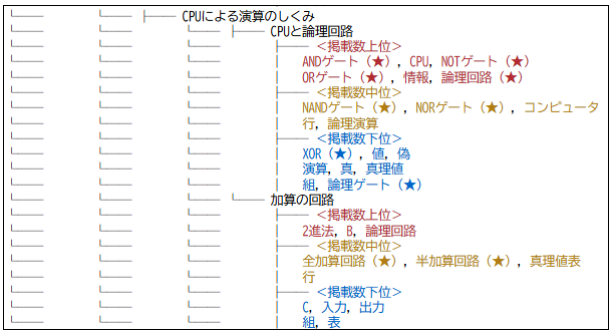


図 8: 「情報 I」における階層構造と掲載状況の可視化（一部抜粋）

との対応関係を体系的に整理し，オントロジーを活用した比較分析を進める予定である．とくに，章や節といった表層的な構成の差異に左右されず，各教科書の内容を比較できるような分析を目指す．

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），2018.

[2] 大学入試センター，令和 9 年度大学入学者選抜に係る大学入学共通テスト出題教科・科目の出題方法等，2025

[3] 赤澤紀子，赤池英夫，柴田雄登，角田博保，中山泰一，情報教科書に現れる用語の変遷：情報 ABC から情報 I・II まで，情報処理学会論文誌，教育とコンピュータ，Vol.10，No.1，pp.13-24，2024.

[4] 中園長新，高等学校「情報 I」教科書の索引に掲載された語句の傾向，2022PC カンファレンス，2022.

[5] 菊谷和也，笹嶋宗彦，高等学校情報科「情報 I」の教科書ごとの重要語の分析と比較検討，人工知能学会全国大会論文集，第 38 回 (2024)，pp.1-4，2024.

[6] 情報処理学会，情報科全教科書用語リスト，2024

[7] 東京都教育委員会，令和 4 年度使用 都立高等学校及び都立中等教育学校（後期課程）用教科書教科別採択結果（教科書別学校数），2021.

[8] 荻谷昌己，最新情報 I，実教出版，2022.

[9] 菊谷和也，笹嶋宗彦，高等学校情報科「情報 I」における教科書からの階層関係抽出法の検討，ARG Web インテリジェンスとインタラクション研究会，第 20 回研究会予稿集，pp.151-154，2024.

[10] 荻谷昌己，高校情報 I Python，実教出版，2022.

[11] 坂村健，高等学校情報 I，数研出版，2022.

[12] 赤堀侃司，東原義訓，坂元章，新編情報 I，東京書籍，2022.

[13] 黒上晴夫，坂田龍也，村井純，情報 I，日本文教出版，2022.

LLMによる要約・詳細化過程におけるハルシネーションの分析

山田 夏稀[1]　安尾 萌[2]*　松下 光範[3]　Junjie Shan[1]　西原 陽子[1]
Natsuki Yamada[1]　Megumi Yasuo[2]　Mitsunori Matsushita[3]　Junjie Shan[1]　Yoko Nishihara[1]

[1] 立命館大学情報理工学部
[1] College of Information Science and Engineering, Ritsumeikan University
[2] 立命館グローバル・イノベーション研究機構
[2] Ritsumeikan Global Innovation Research Organization
[3] 関西大学総合情報学部
[3] Faculty of Informatics, Kansai University

Abstract: 本研究の目的は，大規模言語モデル（以下，LLM）が要約と詳細化を繰り返す過程で情報がどのように変化し，ハルシネーションが生じるかを明らかにすることである．ニュース記事を用い，LLMによる要約・詳細化を複数回繰り返し，出力結果をハルシネーションのタイプ別に分類，原文および出力結果の修辞構造の観点から分析した．その結果，具体的な数値や根拠情報が失われるなど，既存研究で報告されている人のうわさの伝搬行為と同様の変化が確認された．また修辞構造の観点では，LLMがハルシネーションを含めた上で，元となる記事と類似した文章構成のテキストを生成することが確認された．

1　はじめに

Web上の偽情報や誤情報の蔓延は，社会的な分断や意思決定の誤りを引き起こす深刻な問題となっている．誰もが簡単に情報発信が可能な現代において，その拡散速度と影響範囲は増大し続けている．こうした偽情報や誤情報は，発信者個人の悪意や誤解のよって生じるものに限るものではなく，情報の伝搬プロセスの過程でも生じえる．情報が伝播するプロセスの一例として伝言ゲームがある．伝言ゲームは情報が人から人に経由していくことで，最初の情報が変化・劣化する様子を楽しむゲームである．伝言ゲームのプロセスは，もととなる情報の伝達と，伝達の際に抜け落ちた情報の補完の繰り返しとして捉えることができ，実際のWeb上での情報の拡散プロセスの例として用いられることがしばしばある．

LLMの登場により，情報の生成や要約が自動かつ高速に実行可能となった．実際に，LLMを用いてWeb上のニュース記事を要約し，報道内容の概略を投稿するボットなどがソーシャルメディア上で運用されている．しかし，LLMもまた事実と異なる情報を生成する現象（以下，ハルシネーション）を引き起こし，誤った情報を拡散させる懸念がある．Webクローラによる高速な情報収集と，LLMによる高速な情報の生成および要約が組み合わさることで，LLMを介した伝言ゲームは，従来とは比較にならない速度と規模で誤情報を生成・拡散させるという新たな脅威となりうる．

LLMの誤情報拡散は，もととなったニュース記事からLLMが内容を要約し，要約された内容から元の内容が補完される，という処理の繰り返しの過程でハルシネーションが混入し，それが拡散されることで発生すると考えられる．本研究では，LLMを介した伝言ゲームのプロセスを「要約」と「詳細化」の2つの処理の繰り返しと捉える．従来の誤情報研究は，人間の認知バイアスや社会的拡散パターンを中心としてきた[1]．一方で，LLMによって生成される誤情報の段階的な変化に焦点を当てた研究は少なく，要約と詳細化のサイクルを繰り返す中で，元の情報が具体的にどのようなメカニズムで変化し，どの段階でハルシネーションへと劣化していくのか，そのプロセスは解明されていない．本稿では，LLMに要約と詳細化からなる伝言ゲームのプロセスを実行させ，その過程で観察されるハルシネーションの特徴を分析する．

2　関連研究

2.1　LLMとハルシネーションに関する先行研究

Huangらは，ハルシネーションに関する包括的なサーベイを提供し，ハルシネーションが要約タスクにおいて

*連絡先：立命館大学情報理工学部
〒 567-8570 大阪府茨木市岩倉町 2-150
E-mail:{yasuo-ri,nisihara}@fc.ritsumei.ac.jp

も深刻な影響を及ぼすことを報告している [2]．Maynez らは，大規模な人手評価を行い，当時の最先端の要約モデルであっても，生成された要約の多くが入力記事に忠実でない情報を含んでいることを明らかにした [3]．Kalai らはハルシネーションの発生原因について，モデルが不確実な場合に推測で回答しても，人手評価や自動指標によって報酬が与えられる設計になっていることが，その一因であると指摘している [4] ．

2.2 人間社会における伝言ゲームでの誤情報伝搬

情報の伝播が人々の行動に与える影響は，古くから社会の重要な関心事であった．Allport らは噂の伝達実験を行い，伝達する過程で内容が短く平易になる「平均化」，特定の要素が強調される「強調化」，人の既存の知識や信念に沿うように内容が歪められる「同化」，情報に要素や詳細が追加される「付加化」といった現象が起きることを実証した [5]．このような人間の認知的バイアスによる情報変化の知見は，災害時などの現実社会での情報伝播の分析にも応用されている．小笠原らは，災害時の情報伝播に関する研究において，人々が曖昧な状況に対して独自の解釈を追加することで不安を反映した誤情報の種が生まれるプロセスを示した [6]．誤情報が社会パニックを引き起こした実証事例として，有馬らは 1973 年に発生した豊川信用金庫の取り付け騒ぎを詳細に分析している．この事例は，誤情報が社会パニックへと発展するプロセスを記録している．この調査では，豊川信用金庫に就職が決まった友人に対して発された「信用金庫は危ない」という発言が伝播する過程で，主語が曖昧な「信用金庫」から「豊川信用金庫」へと対象が特定化され， さらに，本来の意図とは異なる「経営が危ないらしい」という推量の噂へと意味内容が変化し，最終的には潰れるという根拠のない断定へとエスカレーションしたというプロセスが示されている [7]．これらの先行研究から，情報は伝達過程において，不確実な状況に関する解釈が付け加えられることで誤情報が増加し，社会的なパニックを引き起こしうることが示されている．

2.3 本研究の位置づけ

LLM による情報の反復的な処理は，人間が情報を伝達する伝言ゲームのプロセスと構造的な類似性を持つ．たとえば，ある事象についての内容を簡潔に伝える処理，および要約によって省略された内容を復元する処理を繰り返すことは，あるニュースのタイトルや趣旨のみを伝達し，その内容について事後的に肉付けされるという構造と類似する．LLM の情報処理プロセスが社会に与える影響は増大しつつあることから，豊川信用金庫の事例 [7] や災害時における誤情報伝播で観察される同化・付加化 [6] といった，情報伝達の変化が，LLM においても再現されるのか，あるいは LLM 固有の全く異なるエラーパターンが出現するのかを明らかにすることは，LLM が社会に及ぼす影響を予測する上で不可欠な研究課題である．

以上を踏まえ本研究では，LLM の反復的な情報処理を，人間の伝言ゲームにおける情報の変化と対応させて分析する．内容を要約するプロセスと，元の要約を復元するプロセスを反復するシミュレーションを行い，単一のエラーを特定するだけでなく，サイクルを通じてハルシネーションが蓄積，変化するプロセスを分析することで，LLM が情報の再生成を繰り返す中で，どのような変化パターンを示すかを定性的に解明する．

3 分析方法

3.1 LLM を用いたハルシネーションを含むデータの生成

本研究では LLM が生成する可能性があるハルシネーションを分析するため，人間社会の噂の伝播モデル [5] から，伝達する過程で内容を短縮するプロセスと，短縮された内容を再度復元するプロセスに着目し，LLM が生成するハルシネーションを，この 2 つのプロセスからなると仮定した．本稿では各プロセスをそれぞれ「要約」および「詳細化」と記述する．この 2 つのプロセスを LLM に反復的に実行させることで，ハルシネーションを含む情報を生成させる．

LLM を用いてハルシネーションを含むデータを生成させる手順について説明する．

1. **要約ステップ**：原文，または 1 つ前の詳細化で得られたテキストに対し，LLM を用いて要約をする．要約の文字数は元のテキストの 50%程度と指定した．
2. **詳細化ステップ**：要約ステップで得られたテキストに対し，同じく LLM を用いて詳細化を行う．詳細化の文字数は原文テキストと同程度と指定した．

この調査では，上記の要約ステップと詳細化ステップを各 10 回繰り返した．結果として，原文テキスト 1 件に対し，要約テキストが 10 件，詳細化テキストが 10 件得られた．本論文で使用した LLM は GPT-4（gpt-4-0613）であった．

原文テキストは，Yahoo!ニュースに掲載されたニュース記事を対象とした．Yahoo!ニュースはニュース記事のカテゴリを国内，国際，経済，エンタメ，スポーツ，

IT，科学，ライフ，地域に分類している．この実験ではYahoo!ニュースにおけるニュース記事のカテゴリ分類に基づき，9つのカテゴリからそれぞれニュース記事を1件ずつ取得し，分析に用いた．記事の取得日は2025年4月8日，4月12日，および5月29日であった．

3.2 ハルシネーションの種類分析

生成された詳細化テキストに対し，ハルシネーションの種類を分析する．本研究でのハルシネーションの定義は,「原文テキストに存在しない，あるいは文脈と矛盾する内容が生成される現象」とする．ハルシネーションの内容は既存研究を参考にし，以下の5種類とする [2].

1. **外在的事実誤り**：訓練データに含まれない事実誤り.（ex. 原文にない誤った年号や固有名詞の生成）
2. **内在的事実誤り**：学習に用いたデータには存在する情報だが，現実とは異なる事実を出力する誤り.
3. **指示不一致**：「詳細化する」という指示に対し，原文にはなかった独自の解釈や評価（ex. これは世界経済の健全さを示している）を生成する誤り.
4. **文脈矛盾**：会話履歴や前段の生成内容と矛盾する誤り.
5. **論理的不整合**：生成された詳細化文の内部で，論理的な矛盾（ex. 前半と後半で主張が異なるなど）を引き起こす誤り.

内容分析では，詳細化テキストを1文ずつ人手により評価する．ある文が以下の3つの場合分けのいずれかに該当する場合は，ハルシネーションが含まれる文が生成されたと評価する.

1. 原文テキストに含まれていない情報を含む.
2. 原文テキストの内容と矛盾する記述を含む.
3. 原文テキストの文脈からは支持されない事実に基づかない推論や断定を含む.

ハルシネーションを含む文に対し，先に示した5種類のラベルのいずれかを付与する．ラベルは複数の付与を可能とした．LLMがもととなる記事からハルシネーションを生成する際の変化を観察するため，10回実施された要約–詳細化ステップのうち，初期段階である1回目と2回目のステップで生成された詳細化テキストを対象として内容分析を行う.

3.3 修辞構造タグを用いたハルシネーションの構造分析

LLMにより要約と詳細化が繰り返されることで生成されるテキストについて，構造レベルでのハルシネーションを分析するために，修辞構造タグを用いて分析を行う．修辞構造タグを用いることでテキストの構造が把握できるため，LLMが要約と詳細化を行う際に原文テキストの構造を認識しているのか，また構造を保持した上で生成を行っているのかを確認する．構造変化が原文の持つ固有の特性に強く依存するのであれば,類似したジャンル同士はその変化のパターンもまた類似するという仮説を立てた.

分析では，詳細化された1回目から10回目までの10種類のテキストに対し，テキストに含まれる各文に対し，修辞構造を示すタグを付与した．修辞構造を示すタグは既存研究に基づき [8],「原因」,「条件」,「否定条件」,「目的」,「譲歩」,「対比」,「例外」,「類似」,「代替」,「連言」,「選言」,「例示」,「詳細化」,「言い換え」,「同時性」,「非同時性」,「展開」,「評価」の18種類のタグを使用した.

タグの付与手順はLLMを用いて以下の手順で行った．まず，タグの定義と例をプロンプトとして用意し，詳細化テキストの文と合わせてLLMに入力する．LLMが付与したタグについて，第一著者が原文の文脈と照らし合わせ，修正を行った上でタグ付与を完成させた．使用したLLMはGPT-4（gpt-4-0613）であった.

続いて，付与されたタグを詳細化のテキストごとにベクトル化する．ベクトルの各要素を修辞構造を表すタグとし，テキストに含まれる修辞構造のタグの割合をベクトルの値とする．1つの詳細化テキストに対し,18次元のベクトルが得られる．1つのジャンルごとに18次元のベクトルを，1回目から10回目までの詳細化テキストに付与されたタグの割合をステップ順に連結し，$18 \times 10 = 180$ 次元のベクトルを得る.

続いて，ベクトル間の類似度を算出する．異なるジャンルの2つのベクトルの類似度をコサイン類似度により算出する．これにより，9ジャンル，$9 \times 8/2 = 36$ 個のコサイン類似度が得られる.

4 分析結果

4.1 ハルシネーション内容分析の結果

ハルシネーションの内容分析の結果を表1に示す．内容分析の結果,「指示不一致」,「外在的事実誤り」,「内在的事実誤り」,「文脈矛盾」,「指示不一致+外在的事実誤り」,「指示不一致+内在的事実誤り」,「指示不一致+文脈矛盾」，の7種類のタグとその組合せが観察された．最も多く観察されたハルシネーションは「指示不

表 1: ハルシネーション分類の集計

行ラベル	IT	エンタメ	スポ	ライフ	化学	経済	国際	国内	地域	総計
指示不一致	3	5	2	4	9	3	2	5	8	**41**
外在的事実誤り	0	1	0	0	0	0	2	1	0	4
内在的事実誤り	1	0	2	4	1	1	7	1	0	**17**
文脈矛盾	0	0	0	0	0	0	0	0	1	1
指示不一致＋外在的事実誤り	0	2	0	0	0	0	0	0	0	2
指示不一致＋内在的事実誤り	0	0	0	0	0	1	1	0	0	2
指示不一致＋文脈矛盾	0	0	1	0	0	0	0	0	0	1
総計	4	8	5	8	10	5	12	7	9	68

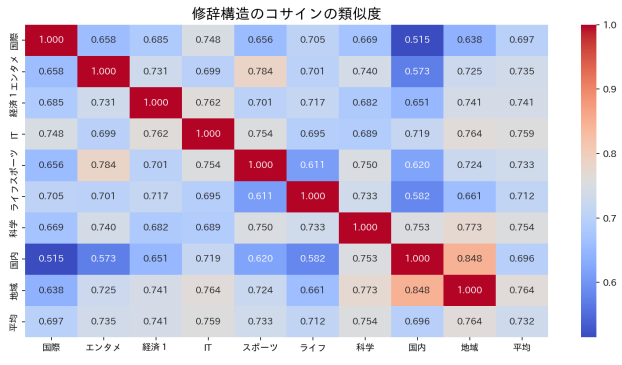


図 1: 修辞構造タグの推移を表すベクトルの類似度を示したヒートマップ．類似度はコサイン類似度．ジャンルごとにベクトルを作成し，ジャンル間の類似度を算出した．

一致」であった．続いて「内在的事実誤り」のハルシネーションが多くなった．一方,「論理的不整合」に該当するハルシネーションは観察されなかった．この原因として，今回の調査において入力されたニュース記事が LLM のコンテキストウィンドウの範囲内であり，一貫性のある文章が生成可能な状況であったこと，また LLM が文章生成において文法や意味的なつながりを優先するため，文内部で矛盾を回避する傾向にあるためだと推測できる．

4.2 修辞構造タグを用いた分析の結果

修辞構造タグを用いた分析の結果として，ジャンル間のコサイン類似度をヒートマップとして図 1 に示す．類似度が高かったジャンルの組合せは，国内と地域（0.848），科学と地域（0.773），IT と地域（0.764）であった．類似度が低かった組合せは，国際と国内（0.515），エンタメと国内（0.573），ライフと国内（0.582）であった．また，IT，地域，エンタメ，科学は他ジャンルとの平均的な類似度が高かった．

表 2: 詳細化テキストに含まれたハルシネーションの例．指示不一致の内，原文テキストにない情報の付加の例．

原文テキスト	詳細化テキスト
ラッパーの Awich がサプライズで登場し、和と融合したラップで盛り上げたり、歌舞伎役者・尾上菊之助らが舞を披露したりするなど、日本色溢れるステージで華を添えた。	また、コブクロや絢香、imase、尾上菊之助らも登場し、歌や演技で万博の開幕を祝いました。
シンガーソングライター・imase とともに、最新のテクノロジーと音楽が融合したステージで盛り上げ、会場の心をひとつにした。	これらの出演者は、日本の伝統文化と最新技術を融合させたパフォーマンスを披露し、**万博のテーマ「共生・創造・未来」を象徴する華やかな内容となりました。**
吉村洋文大阪府知事、同博覧会名誉会長の石破茂首相らがあいさつ。	式典では、石破茂首相や吉村洋文大阪府知事も挨拶を行い、**日本の進歩的な技術や文化を世界に向けて発信する決意を表明しました。**

5 考察

5.1 ハルシネーションの内容分析に関する考察

表 1 において，最も観察された回数が多かったのは「指示不一致」のハルシネーションであった．指示不一致の中でも頻繁に見られたのは，原文テキストにない情報の追加であった．この例を表 2 に示す．表 2 では，原文テキストは万国博覧会の開会式に関する記事であった．1 回目の詳細化において，原文テキストには含まれないが，万国博覧会のテーマに関する記述が追加された（表 2 中太字箇所）．追加された原因としては，LLM が詳細化を実行する際に原文テキストの情報を忠実に保持することにより，より流暢で尤もらしいテキストの生成を優先した結果と考えられる．また，このハルシネーションは，人間による噂伝播モデルで指摘される同化および付加化のプロセスとも類似している．具

体的には，原文テキストのトピックから連想される情報の追加（万国博覧会のテーマ追加）は，読み手の既存知識に沿うように内容が歪められる同化と類似していると考えることができる．

さらに，原文テキストに含まれている重要な情報，特定の数値データや統計情報などが，詳細化のテキストでは失われることが多かった．他には，原文テキストで記述されていた事実や発話の内容を変更する現象も確認された．スポーツの監督や選手によるインタビュー記事において，その発言のニュアンスや内容が，原文テキストの内容を忠実に復元せず，異なる形に置き換えられていた．

5.2 修辞構造タグを用いた分析結果に関する考察

ジャンル間の修辞構造タグの類似度について，類似度のスコアが高かった組合せは，国内と地域，経済とエンタメ，IT と地域であった．これらの組合せに対し，原文テキストを段落単位に分割し，全体の構成を定性的に比較したところ，原文テキストの文章構成と論理展開がジャンル間で類似する傾向が見られた．表 3 は，国内ジャンルの記事で扱われた関西万博開幕時の喫煙問題と地域ジャンルの記事で扱われた愛知県の特殊詐欺リクルーター摘発事件について，両者の段落構成を比較した例である．両記事は扱うテーマこそ異なるものの，時系列に沿った説明，中盤での前提情報の挿入，関係者のコメント提示，そして最後に背景や動機を述べるといった構造が共通していることがわかる．一方で，要約と詳細化を繰り返す中で，特定の修辞構造の単調増加や単調減少の傾向は見られなかった．

図 2 に，原文と 1 回目から 10 回目までの詳細化で見られた修辞構造タグの出現回数の推移を示したグラフを示す．2 つの図からは，修辞構造タグが要約と詳細化が 1 回行われるごとに，含まれる修辞構造タグの数に変化があることが示されている．これは，LLM が生成するハルシネーションを自然な文脈のテキストに組み込むために，新たな修辞関係を生成した結果と考えられる．実際に，表 4 に原文と詳細化の文を比較し，ほぼ同一の内容で修辞構造が変化した例を示す. この例では，関西万博の開会式において登場したゲストとパフォーマンスについて説明するという同一の内容を記述しているものの，修辞構造が変化していることが確認できる．

ジャンル間で類似度の高い組み合わせがあるという結果は，ジャンル間での修辞構造の変化の仕方が類似していたということを示している．修辞構造の変化の仕方が類似した原因として，原文テキストの文章構成と論理展開が似ていた可能性が考えられる．したがって，LLM による要約と詳細化の繰り返しによって生成された文章は，原文テキストの文章構成に依存して構造が類似することが示唆される．

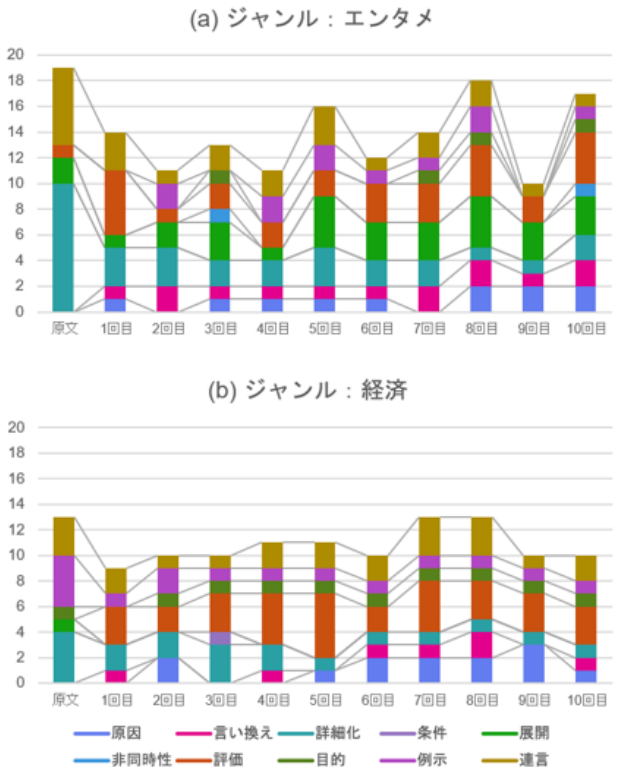


図 2: 修辞構造タグの出現回数の推移

6 おわりに

本研究では，LLM による要約と詳細化が繰り返されることで生成されるハルシネーションの分析を行なった．分析のために，ニュース記事を原文テキストとして，LLM を用いて要約と詳細化を繰り返してハルシネーションが含まれるテキストを生成した．生成されたテキストのうち，詳細化の段階で得られたテキストと原文テキストを比較することにより，ハルシネーションの種類を分析した．

分析の結果，生成されるハルシネーションとしては，(1) 原文テキストにはない情報が追加される，(2) 数値データ，統計データが欠落される，(3) 発言のニュアンスや内容が書き換えられる，などが多いことがわかった．さらに，各テキストに含まれる文に対し，修辞構造の分析を行った結果，ハルシネーションを含めた上で自然な文脈のテキストを生成するために，原文テキストの文章構成を踏まえて，類似した文章構成のテキストを生成することが分かった．

表 3: 国内ジャンルと地域ジャンルにおける段落構成の比較（類似度 = 0.848）

国内	地域	共通する構成特徴
時系列進行，途中で前提（ルール）を挿入，末尾に背景を提示	時系列進行，途中で兆候や動機を補足，末尾に背景・動機を提示	時系列ベース，中盤に前提／兆候，末尾に背景（動機）を配置
第 1 段落：万博開幕と状況説明	第 1 段落：摘発の事実を提示	いずれも事実提示で開始
第 2–3 段落：喫煙行為と「全面禁煙」という前提提示	第 2–3 段落：不審行動と，それを察知できた理由（兆候）	中盤で前提や兆候を示す構成
第 4–6 段落：違反状況，違反者コメント，運営側対応	第 4–8 段落：母親の行動，警察対応，供述，事件詳細	中盤〜後半で行動・供述の詳細を提示
第 7 段落：過去事故への言及（背景）	第 9 段落：母親の動機と警察の呼びかけ	最終段落に背景・動機を配置

表 4: 詳細化による修辞構造の変化例（情報の分割と評価の付与）

原文テキスト	詳細化テキスト
ラッパーの Awich がサプライズで登場し、和と融合したラップで盛り上げたり、歌舞伎役者・尾上菊之助らが舞を披露したりするなど、日本色溢れるステージで華を添えた。 **タグ：**詳細化， 評価	サプライズゲストとして、ラッパーの Awich が登場し、そのパワフルなパフォーマンスが会場を盛り上げました。 **さらに**、歌舞伎役者の尾上菊之助らが**美しい**舞を披露し、伝統的な日本の芸能が見られるなど、日本色溢れるステージが繰り広げられました。 **タグ：**連言， 評価

今後は，LLM ごと，あるいはニュース記事ジャンルごとに追加の分析を行い，結果を比較することで，発生するハルシネーションの種類とその伝達プロセスの違いについて明らかにする．また，修辞構造の順序や構造について追加の分析を行い，構造レベルでの情報の変化を明らかにする．

参考文献

[1] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. The science of fake news. *Science*, 359(6380):1094–1096, 2018.

[2] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions, 2023.

[3] Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. On faithfulness and factuality in abstractive summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1906–1919, 2020.

[4] Adam Tauman Kalai, Ofir Nachum, Santosh S. Vempala, and Edwin Zhang. Why language models hallucinate, 2025.

[5] Gordon W. Allport and Leo Postman. *The Psychology of Rumor*. Henry Holt and Company, 1947.

[6] 小笠原 盛浩, 川島 浩誉, and 藤代 裕之. マスメディア報道は Twitter 上の災害時流言を抑制できたか？—2011 年東日本大震災におけるコスモ石油流言の定性的分析. **関西大学社会学部紀要**, 49(2):121–140, 2018.

[7] 有馬 守康, 齋藤 哲哉, 小林 創, and 稲葉 大. 「取り付け騒ぎ」に関する理論的・実験的分析と事例との整合性に関する考察. **日本大学経済学部経済科学研究所紀要**, 49:45–53, 2019.

[8] Harry Bunt, Volha Petukhova, Emer Gilmartin, Catherine Pelachaud, Alex Fang, Simon Keizer, and Laurent Prévot. The ISO standard for dialogue act annotation, second edition. In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, pages 549–558, 2020.

PPDACサイクルオントロジーに基づく「総合的な探究の時間」の指導計画作成補助システムの試作

Prototyping of the Instructional Planning Support System for“Period for Inquiry-Based Cross-Disciplinary Study”Based on the PPDAC Cycle Ontology

* 堀之内逸人[1] 林宏樹[2] 笹嶋宗彦[1]
Hayato Horinouchi[1] Hiroki Hayashi[2] Munehiko Sasajima[1]

[1] 兵庫県立大学
[1] University of Hyogo
[2] 雲雀丘学園中学校・高等学校
[2] HibarigaokaGakuen Junior&High School

Abstract: One challenge in teaching Period for Inquiry-Based Cross-Disciplinary Study (hereafter referred to as“inquiry-based learning”) is the difficulty experienced by teachers with limited experience in this area when creating lesson plans.In this study, we developed a prototype system to assist inexperienced teachers in creating instructional plans for inquiry-based learning, referencing the PPDAC cycle ontology as a guidance model for inquiry-based learning.This system is an instructional planning support system that outputs the proposed instructional plan and Examples of learning content generated by AI, based on the input of total class hours and scope of instruction, research themes.

1 はじめに

日本の高等学校では，2022 年度から平成 30 年告示高等学校学習指導要領 [1]（以下,「新学習指導要領」）が年次進行で実施が始まった．新学習指導要領において,「総合的な探究の時間」は「総合的な学習の時間」から名称を変更し，全ての生徒が履修する授業となった．総合的な探究の時間は，探究学習を行う授業である．

総合的な探究の時間の課題としては，専門的な教員免許が存在せず，他教科の教員免許を持った教員が授業を担当しており，文部科学省検定済教科書もないことが挙げられる．この点について稲永 [2] は，探究学習の指導には十分な学術経験が必要であり，修士レベルの学術経験がないと苦慮すると指摘している．文部科学省の調査よれば，現役の公立高等学校教員の学歴構成は，大学院修了者の割合が 16.5 %である [3]．このことから，多くの高等学校教員にとって探究学習の指導には難しさがあると想像できる．

以上の背景から，多くの教員は総合的な探究の時間に対して負担と不安を抱えていると考えられる．高等学校理科教員を対象にした調査によると，探究活動を導入する上での課題として，探究活動を計画するために多くの時間が必要であることを 9 割以上の教員が挙げたと報告されている [4]．さらに，他の調査では，探究学習の指導に対して，探究の過程への指導方法の難しさや指導計画の作成の難しさに不安を抱える教員が多いと報告されている [5][6]．これらのことから，探究学習を担当する教員，特に探究学習の指導経験が浅い教員にとって，指導内容や指導計画を立てることは困難であると考えられる．これら教員の支援をするための枠組みが必要である．

本研究では，指導経験の浅い教員が総合的な探究の時間の指導計画を作成することを補助するシステムの実現を目標として，探究学習の指導モデルオントロジーを参照して指導計画案を作成するシステムの試作を行った．本システムは，授業時間数と授業範囲，探究テーマを入力すると，指導計画案と生成 AI が作成した学習内容の事例を出力する，指導計画作成補助システムである．本システムが参照する探究学習の指導モデルとして，PPDAC サイクルオントロジー [7] を採用し，先行研究で林らが作成した PPDAC サイクルオントロジーに指導計画案の作成に必要な知識を追加した．オントロジーを用いることで，暗黙的な探究学習の指導の流れや生徒の学習状態，学習内容といった情報を提示できる．指導計画案は，PPDAC サイクルオントロジーに

*連絡先：兵庫県立大学社会情報科学部社会情報科学科
〒 651-2197 兵庫県神戸市西区学園西町 8 丁目 2-1
E-mail:ad25p065@guh.u-hyogo.ac.jp

定義されている指導の流れに基づき，学習内容や指導内容，推定された時間配分を要素として構成された表形式で出力される．本システムを用いることで，指導経験の浅い教員が指導計画を作成することに関する負担の軽減に期待できると考えている．

提案するシステムの利点としては，出力する指導計画案は指導モデルオントロジーに沿うものであり，既存の指導モデルの知識について議論を行い変更が生じた場合は，オントロジーの編集を行うだけでよく，インタフェース部プログラムのアルゴリズムを変える必要が無いことが挙げられる．

2 PPDACサイクルオントロジー

本研究では，探究学習の指導モデルとしてPPDACサイクルオントロジー [7] を用いることにした．

オントロジーとは，対象の現実世界に関する概念とそれらの関係性について記述したものである．オントロジーは，共通語彙の提供や暗黙情報の明示化，知識の体系化といった機能を持っており，知識の共有や再利用性を向上することができる [8][9].

PPDACサイクルオントロジーとは，問題解決プロセスであるPPDACサイクルを活用した探究学習の指導方法をモデル化したものである [7]．日本の高等学校における探究学習の問題解決プロセスに関する知識を体系化し，体系化した知識をPPDACサイクルに基づいて構造化している．PPDACサイクルとは，Problem（問題），Plan（計画），Data（データ），Analysis（分析），Conclusion（結論）の５つのフェーズから構成される，データを利用した問題解決の手法である．

PPDACサイクルオントロジーには，生徒の探究学習の学習状態を表現するための知識を体系化した部分（以下,「学習状態体系化部分」）と学習状態体系化部分で定義した学習状態と状態間の関係についての知識をPPDACサイクルへ構造化した部分（以下,「PPDAC構造化部分」）がある．学習状態体系化部分には，学習状態と，状態間の移行の仕方と，移行の条件が定義されている．PPDAC構造化部分では，学習状態体系化部分を構成する概念定義をPPDACサイクルに構造化することで，問題解決プロセスと生徒の学習状態の移行の仕方の関係が明確化されている．PPDACサイクルオントロジーは，研究の仕方についての専門家であるベテランの高校教員と大学教員から，探究学習の手順を明示化出来ていると，一定の評価を受けている [7]．このことから，本研究ではPPDACサイクルオントロジーを指導モデルとして採用した．

3 オントロジーの追加構築

林らが構築したPPDACサイクルオントロジーに，指導計画案の作成に必要な知識を追加構築した．

本システムが出力する指導計画案は,「フェーズ名」,「学習内容」,「学習内容説明」,「学習手法」,「時数」の5つの要素で構成される．指導計画案の各構成要素はPPDACサイクルオントロジーから情報を取得する．「フェーズ名」はPPDACサイクルの各フェーズ,「学習内容」は，生徒が行う学習の内容である．生徒の学習状態の変化を学習とすると，状態の移行の条件を学習すべき項目と捉え，プロパティに設定している内容が指導計画案における「学習内容」にあたると考えた．

「学習内容説明」,「学習手法」,「時数」は，既存のPPDACサイクルオントロジーには定義されていないため，新たに知識を追加した．追加構築したオントロジーは，オントロジー構築・利用環境の「法造」上で実装した [10]．追加構築したPPDACサイクルオントロジーの学習状態体系化部分の一部を図1，PPDAC構造化部分の一部を図2に示す．

「学習内容説明」とは，ユーザが理解できるために学習内容を解釈した説明文である．学習内容は，要約された文言になっておりこのままでは理解が難しいため，学習内容にあたるプロパティを設定した経緯をたどり説明文を作成した．例として，学習内容「予想した解の有無」は，学習内容説明を「今回の探究学習において明らかにさせたいこと（問いに対して予想した解）は何かを考え，仮説とする.」とした．同様に，学習内容にあたるすべてのプロパティにおいて説明文を作成し，法造の定義ペインの内容説明欄に記述した．

「学習手法」とは，学習内容に取り組むための手法の例を示したものである．学習内容をさらに具体化させる目的で追加した．文部科学省が発表している「今，求められる力を高める総合的な探究の時間の展開」[11], 探究学習の事例が紹介されている書籍「『探究』学習図鑑」[12], 情報Iの教科書 [13] [14] [15] [16] に掲載されている事例を該当する学習内容に当てはめた．また，一部に「発散的思考」「収束的思考」といった思考手法の分類を導入した．例として，学習内容「理想状態調査」には，学習手法「資料を比較する」,「グラフを読み解く」,「発散的思考」,「収束的思考」とした．学習内容のスロットとしてオントロジーに記述した．

「時数」とは，指導計画案において各学習内容を実施するために必要と見積もられる時間を数値で表したものである．PPDACサイクルオントロジーには，フェーズごとに時間配分を定義した．時数推定のため，神奈川県立多摩高等学校の学校設定教科「Meraki」の年間指導計画 [17] からPPDACサイクルの各フェーズの時数を調査した．各フェーズの時間配分を推定した結果,（Problem：Plan：Data：Analysis：Conclusion）＝（15.45：

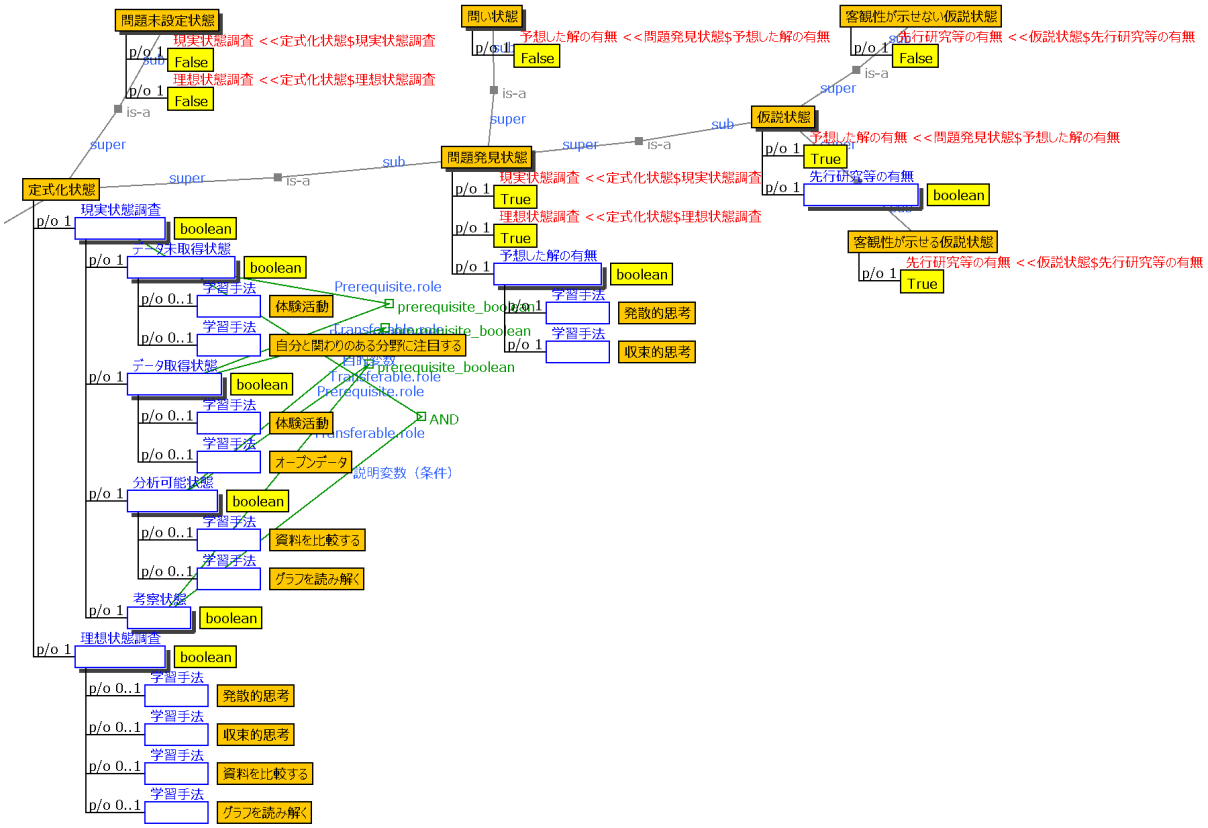


図 1: 学習状態体系化部分の一部

2.95 : 3.4 : 4.95 : 5.75)（単位は単位時間）となった．時間配分は，オントロジーの PPDAC 構造化部分における各「フェーズ」概念のスロットとして追加構築した．

4　システムの設計と試作

本章では，システムの概要と試作したシステムの動作例を示す．

4.1　システム概要

図 3 に本研究で提案するシステムの概要図を示す．

本システムは，授業時間数と授業範囲，探究テーマを入力すると，指導計画案と生成 AI が作成した学習内容の事例を出力する，探究学習の指導計画作成補助システムである．想定するユーザは，探究学習の指導経験が浅い教員である．システムが出力する情報は，指導モデルが定義された PPDAC サイクルオントロジーに基づくものである．オントロジー解析機能において，オントロジーを XML ファイルの形式で読み込み，解析を行い，指導の流れや指導内容といった情報を取得する．指導計画案作成機能では，オントロジーから取得した情報を基に時間配分を計算し，指導計画案を作成する．探究テーマが入力された場合は，生成 AI が探究テーマに即した学習内容の事例を作成する．出力機能によって，ユーザに作成された指導計画案と学習内容の事例を提示する．

4.2　システムの動作例

本システムは，WEB アプリケーションとして開発試作を行った．開発言語は，Python, HTML / CSS, JavaScript である．WEB アプリケーション作成のため，Python のフレームワークの Flask を用いた．

4.2.1　オントロジー解析機能

システムを起動すると，オントロジー解析機能が実行される．オントロジー解析機能では，参照する指導モデルオントロジーである PPDAC サイクルオントロジーを XML ファイルで読み込み，解析する．まず，PPDAC 構造化部分から各フェーズに対して初期状態，中間状態，終了状態の順番に学習状態を取得する．これらの学習状態が，PPDAC サイクルにおいて移行すべき学習状態にあたる．次に，取得した学習状態を学習状態体系化部分と照合し，該当する学習状態のプロパティが boolean または False の場合に，学習内容と学習内容説明，学習手法を取得する．学習状態を取得した順番で，学習状態と学習内容に付随する情報を表形式にする．後

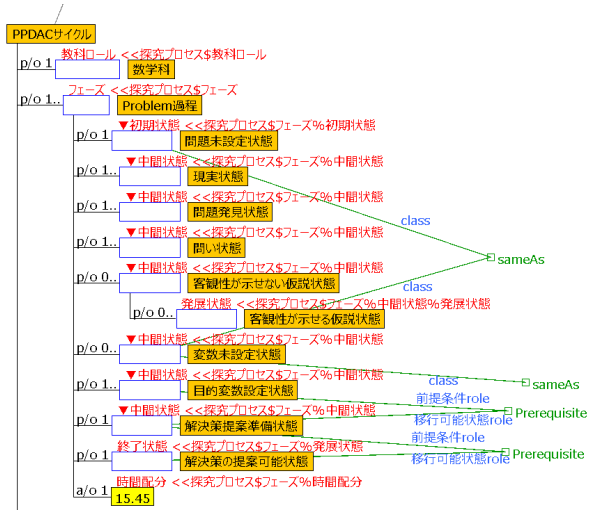


図 2: PPDAC 構造化部分の一部

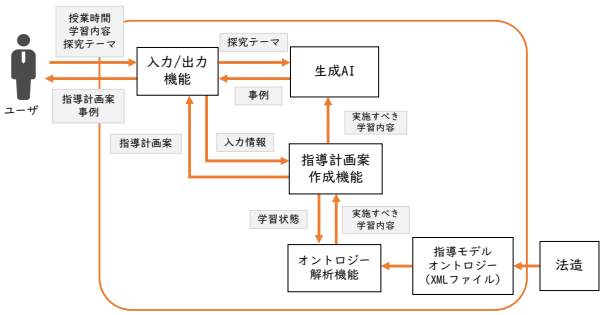


図 3: システム概要図

の機能でオントロジーに基づく必要がある場合は，この機能で解析された情報を利用する．

4.2.2 システム入力画面

図 4 に入力画面を示す．入力画面はシステム起動時に表示される．授業時間数と授業範囲，探究テーマを入力する画面である．授業時間数欄には，探究学習に充てる単位時間数を数値で入力する．授業範囲の入力欄は，指導計画を立てようとする探究学習の段階に対応する，PPDAC サイクルのフェーズをチェックボックスにて入力する形式である．探究テーマ欄は，探究学習の学習内容に即した学習活動を教員が具体的に検討するための事例を確認することを目的とする場合に入力する．入力が完了したら,「次へ」ボタンを押すと入力情報を基に指導計画案が作成される．

4.2.3 指導計画案作成機能

システムへの入力後，指導計画案作成機能が実行される．まず，オントロジー解析機能で作成された表から授業範囲に該当するフェーズ部分を抽出する．次に，

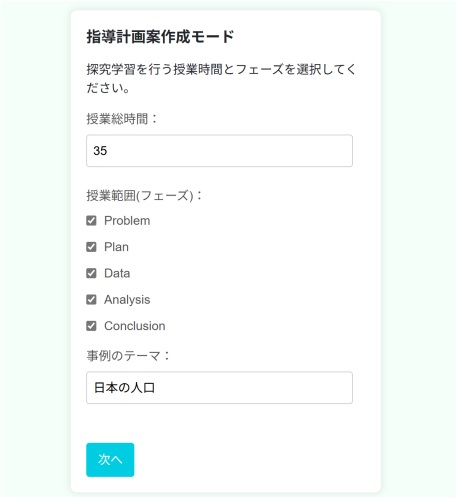


図 4: システム入力画面の例

入力された授業時間数に対し，オントロジーに定義されている時間配分に基づいて学習内容ごとの時数を推定する．時数推定後,「フェーズ名」,「学習内容」,「学習内容説明」,「学習手法」,「時数」の 5 つの要素で構成される指導計画案を作成する．

4.2.4 指導計画案出力画面

図 5 に指導計画案出力画面を示す．画面上部には，各フェーズの時間配分の目安と各フェーズ内の学習内容の数を表示する．各フェーズ内の学習内容の数を示すことで，例として「課題発見」フェーズにおいてどの程度の取り組みが必要とされるかを確認することができる．画面中央には，作成された指導計画案を表示する．指導計画案は，表部分をスクロールすることで下まで閲覧することができる．この指導計画案によって，ユーザは探究学習の全体の指導の流れを確認することができる．

指導計画案の任意の行を選択すると，その行の学習内容についての情報を提示する詳細情報出力画面に遷移する．

4.2.5 詳細情報出力画面

図 6 に詳細情報出力画面を示す．詳細情報出力画面では，システムが提案した指導内容がユーザにとって分かりにくい場合と，探究学習が通常より早く進んで発展的な学習も可能になった場合のために作成した．指導計画案出力画面で，上のいずれかの理由により，ユーザが指導計画案の任意の行を選択した場合に遷移する．詳細情報出力画面では，選択された行の学習内容についての情報を表示する．学習内容についての情報は，指導計画案の要素を基にした情報である．また，選択された学習状態の発展状態がオントロジーのプロパティに設定されている場合，発展状態へ移行するための学

図 5: 指導計画案出力画面の例

図 6: 詳細情報出力画面の例

習内容についての情報を表示する．さらに，学習手法のリンクをクリックするとその手法の説明を表示するページに遷移する．

4.2.6 生成 AI による事例作成

本システムには，学習内容を具体化する事例を生成 AI により自動生成する機能を備えた．本機能は入力画面で探究テーマが入力された場合に実行される．オントロジーに定義された学習内容に関する文言は一般的な表現であるため，指導経験が浅い教員にとっては理解がしにくい可能性がある．そのため，一般的な学習内容の理解促進を目的として，生成 AI を活用して学習内容の具体的な事例を提示する．生成 AI にはプロンプトとして，オントロジーに定義されている学習内容説明と学習手法に加え，ユーザが入力した探究テーマを与える．生成 AI は，探究テーマに即した学習内容の具体的な事例と学習内容に取り組むための手法を用いた活動の事例を生成する．生成された事例は，図 7 に示すように事例出力画面に出力する．なお，ここで生成される事例はあくまでも学習内容の理解の補助のための参考情報であり，必ずしも出力された事例通りに授業を進めるべきものではない．

本システムの試作において，現時点では Groq 社が提供する API を通じて，OpenAI 社が公開している大規模言語モデル 「GPT-OSS-120B」を用いた．

4.3 指導助言への利用

本システムは，教員が生徒の学習状態の適切な把握と適切な指導助言を行うことを補助する目的とした機能への利用も可能であると考えている．具体的には，学習状況を入力すると現在の学習状態と次に取り組むべき学習内容を出力する機能である．入力において，現在の学習状態に当てはまるフェーズを選択した後，該当のフェーズにおいて取り組むべき学習内容が列挙される．さらにユーザは列挙された学習内容の中から達成していると判断した学習内容をすべて選択することで学習状況を入力する．システムは入力された学習状況から現在の生徒の学習状態を探索し，次に取り組むべき学習内容とともに出力する．現状の課題として，学習内容を表す文言が一般的な表現であることから，指

図 7: 生成 AI 作成事例出力画面の例

導経験が浅い教員にとって，個別の具体的な状況から一般的な学習状況を捉えることや一般的な学習内容から個別の指導助言を行うことへの難しさがあることが挙げられる．

5 おわりに

本論文では，PPDAC サイクルオントロジーに指導内容に関する知識を追加構築し，追加構築されたオントロジーに基づく探究学習の指導計画作成補助システムの試作を行った．このシステムを用いることで，ユーザは探究学習の指導の流れと学習内容の情報を得ることができ，指導経験の浅い教員が探究学習の指導や指導計画を作成することへの負担を減らすことが期待できると考えている．

PPDAC サイクルオントロジーに指導計画案の構成要素である学習内容説明，学習手法，時数の知識を新たに追加構築した．オントロジーを用いることにより，暗黙的な探究学習の指導の流れや体系化された学習状態，学習内容といった情報を提示できた．さらに，生成 AI を利用することで，オントロジーに基づく一般的な知識を各テーマに適応させた具体的な事例を提示できた．

本システムについては，現時点では評価を実施していないため，今後，評価実験を行う予定である．システムの利用が教員による指導計画の作成に及ぼす効果を定量的に評価するとともに，アンケート調査を通じて定性的な観点からも評価することを検討している．また，実際の高等学校教員に協力していただきながら今後取り組んでいく予定である．

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），2018.
[2] 稲永由紀：「総合的な探究の時間」の指導を支える教員の学術経験：学士課程教育をめぐる状況と教員養成上の課題，Rcus Working Paper, No. 12, pp. 1-9, 2020.
[3] 文部科学省：令和 4 年度学習教員統計調査，2022.
[4] 小坂那緒子，松原憲治：高等学校普通科における探究の過程の実施状況に関する予備的調査，日本科学教育学会研究会研究報告，Vol. 37, No. 6, pp. 123–126, 2023.
[5] 池田政宣，村瀬公胤，武田明典：「総合的な探究の時間」の導入に向けた高等学校教員のニーズ調査，神田外語大学紀要，第 32 号，pp. 451-471, 2020.
[6] 探究学習研究会：「探究学習」とはいうけれど，晃洋書房，2024.
[7] 林宏樹，笹嶋宗彦：日本の高等学校におけるオントロジーを用いた探究学習の知識モデリング PPDAC サイクルオントロジーの構築，人工知能学会論文誌，40 巻，3 号，C-O82_1-17, 2025.
[8] 溝口理一郎，古崎晃司，來村徳信，笹嶋宗彦：オントロジー構築入門，オーム社，2006.
[9] 溝口理一郎：オントロジー研究の基礎と応用，人工知能学会誌，Vol. 14, No. 6, pp. 978-988, 1999.
[10] 古崎晃司，來村徳信，佐野年伸，本松慎一郎，石川誠一，溝口理一郎：オントロジー構築・利用環境「法造」の開発と利用，人工知能学会論文誌，17 巻，4 号，pp. 407-419, 2002.
[11] 文部科学省：今，求められる力を高める総合的な探究の時間の展開，2023.
[12] 田村学，廣瀬志保：高校生のための「探究」学習図鑑，学事出版，2022.
[13] 坂村健，高等学校 情報 I，数研出版，2022.
[14] 赤堀侃司，東原義訓，坂元章，新編情報 I，東京書籍，2022.
[15] 荻谷昌己，高校情報 I Python，実教出版，2022.
[16] 黒上晴夫，坂田龍也，村井純，情報 I，日本文教出版，2022.
[17] 神奈川県立多摩高等学校：令和元年度指定スーパーサイエンスハイスクール研究開発実施報告書（第 5 年次），2024.

対話グラフの話題遷移に基づく対話パターン分析

Analysis of Dialogue Patterns Based on Topic Transitions in Dialogue Graph

野本 匠馬 [1]　赤石美奈 [1,2]

Takuma Nomoto[1], Mina Akaishi[1,2]

[1] 法政大学 大学院 情報科学研究科
[1] Graduate School of Computer and Information Sciences, Hosei University
[2] 法政大学 情報科学部
[2] Faculty of Computer and Information Sciences, Hosei University

Abstract: *In decision-making dialogues, such as everyday conversations or lively discussions, unintended misalignments in understanding or topic shifts can occur between speakers. Such phenomena can become obstacles hindering smooth dialogue.*
Therefore, this research develops a system to visualize the state of dialogue, thereby providing conversational support. The dialogue patterns referred to in this paper encompass topic transitions, coherence, and shifts in leadership during conversation. By making these visible during dialogue, participants can intuitively grasp what is currently being discussed and whether the conversation is aligned, thereby facilitating smoother communication.

1. はじめに

日常会話や活発な議論などの意思決定を目的とした対話において、話者間で意図せずに認識のずれや話題の食い違いが生じることがある。このような現象は、円滑な対話を妨げる障害となる可能性がある。

そこで本研究では、対話の様子を可視化するシステムを開発することで、対話における話題の遷移を可視化し、話題の追従、拡大、乖離といった対話状態の時系列分析を行う。本稿における「対話の様子」とは、対話中の話題遷移や噛み合い方、主導権の遷移を指す。これらを 2 次元空間上の物理的な距離やグラフの形状として提示することで、参加者は複雑な話題の推移や現在の対話の状態を視覚的かつ直感的に把握することが可能となる。これにより、参加者は認識のずれへの気付きや軌道修正を即座に行えるようになり、円滑な対話の支援に繋がる。

本システムの有効性を検証するため、対話が噛み合っている状態とそうでない状態を定量的に判定する指標を提案し、ケーススタディを示した。複数名の被験者にヘッドホンの新規購入をテーマとした対話を行ってもらい、対話データを収集した。

分析の結果、対話が噛み合ったとされる区間では、両者の話題が特定のポイントで収束するパターンが確認できた。 この結果は、本システムが話者間の認識のずれや対話の噛み合い方を客観的に捉える上で有効であることを示唆している。

2. 関連研究

本研究に関連する従来の研究は、「概念空間の可視化支援」「対話の時系列構造分析」「対話の行動・リアルタイム支援」の 3 つに大別できる。本研究の立ち位置を明確にするため、各分野の代表的な研究と本研究との差異を述べる。

2.1. 概念空間の可視化支援

思考や議論の構造化を支援する研究として、角ら[1, 2]はキーワードの関連性を統計的手法で 2 次元空間に自動配置するシステム（CSS, AIDE）を開発した。これにより、議論の全体構造の認識や新たな発想の獲得を支援できることが示された。

しかし、これら従来の手法は、主に静的なテキスト群の構造化や発想支援に焦点を当てている。これに対して本研究は、リアルタイムで進行する対話において、話者間の関心領域が動的にどう遷移し、いかに「認識のずれ」が生じるかという、対話の「プロセス」そのものを可視化する点に違いがある。

2.2. 対話の時系列構造分析

対話の時系列的な流れを分析する研究として、Okada ら[3]は、会議の発散・収束といった状態を定量化し、合意形成の評価と相関があることを示した。

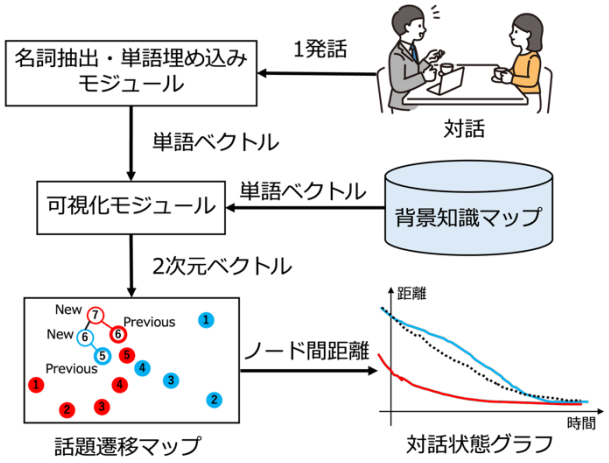


図 1: 対話可視化フレームワークのシステム構成図

しかし、これらの研究は、会議全体の状態評価や、テキストの主題の一貫性診断が主な目的である。これに対して本研究は、話者 A と話者 B それぞれの関心領域がリアルタイムで独立して遷移しているか、あるいは収束しているかという、2 者間の「意味的な関係性」の変化として対話の噛み合い方を分析する点に特化している。

2.3. 対話の行動・リアルタイム支援

対話のリアルタイム支援に関する研究として、柳楽・水本ら[4]は、主観に頼らない客観的な振り返りを支援するため、「総発話時間」や「ターンテイク」といった発話の「量」や「行動」に着目した可視化システムを構築した。

しかし、これらの研究は主に行動情報に基づいており、対話の「質」の側面は考慮していない。行動は活発でも、それぞれの話者が話している話題が意味的に遠くなっているなどの状態は検出できない。これに対して本研究は、自然言語処理を用いて発話された話題に基づき、対話の噛み合い方をグラフ構造の変化として客観的に捉える点に新規性がある。

3. 対話可視化フレームワークの概要

本章では、提案する対話可視化フレームワークの背景とシステム概要について述べる。

3.1 節では、対話の状態を定量的に捉えるための指標である「トピック間距離」の定義と、それに基づいた可視化ツールについて説明する。3.2 節では、本フレームワークの全体的なシステム構成について述べる。

3.1. トピック間距離に基づく対話の可視化

対話中の話題の遷移、噛み合い方を捉え、その状態を定量的に分析するために、本研究では「トピック間距離」を定義する。

まず、対話の基本的な構成要素として、話者が発話した話題を「発話トピック」と定義する。発話トピックは、話者の発話テキストから抽出された名詞を単語ベクトルとして表現することで、計算可能な意味空間上の点として扱われる。

トピック間距離とは、この発話トピックを基に、一話者の発話に出現するトピック間の距離、または他者との対話に出現するトピック間の距離のことである。この距離は、単語ベクトル間の意味的な類似度をユークリッド距離として算出したものであり、この距離を測定することで、対話が意味的にどの状態にあるかを判断することが可能となる。

本研究ではトピック間距離を、分析の目的に応じて以下の 2 種類のトピック距離として定義した。

1. **発話トピック距離**
   発話トピック距離は同一話者の最新の発話トピックと、その直前の発話トピック間の距離として定義される。
   距離が小さい場合はトピックが収束あるいは維持されており、話者の話題が一貫していることを示す。逆に距離が大きい場合はトピックが拡散されたことを示す。
2. **対話トピック距離**
   2 者の話題が意味的にどの程度近いか示すため、対話トピック距離は 2 者の最新の発話トピック間の距離として定義される。
   この距離が一定期間小さい場合には、両者の話題が収束または類似しており、逆に距離が大きい場合は、両者の話題が乖離している。

本研究では、これらのトピック間距離を用いることで、対話の様子の全体像を把握することを目的とする。トピック間距離の絶対的な値と、その時間変化を組み合わせることで、対話が「収束」「乖離」や「追従」「膠着」といった、どのような状態にあるかを判別できる。このようなトピック間距離の変化を直感的に把握するため、その距離を物理距離とし 2 次元空間上に表示したものを話題遷移マップと呼ぶ。また、トピック間距離の時間変化を定量的に把握するため、時系列の折れ線グラフとして描画する対話状態グラフを作成する。これらの可視化ツールを用いることで定量的に対話の状態を判別する仕組みを提案する。

3.2. システム構成

本研究では、対話中の話題遷移、話題の近さ、対話の状態をリアルタイムで可視化・分析できるフレームワークを提案する。

この目的を達成するため、本フレームワークは、

表 1: 形態素解析器の辞書変更による解析結果

解析テキスト	取得名詞（デフォルト）	**取得名詞（NEologD）**
リモートワークで機械学習エンジニアとして働いています。	リモート、ワーク、機械、学習、エンジニア	**リモートワーク、機械学習、エンジニア**
好きなボカロPは米津玄師です。	好き、ボカロ、P、米津、玄、師	**好き、ボカロP、米津玄師**

対話の様子を分析する 2 つの主要な可視化機能として、話題遷移マップと対話状態グラフを提案する。話題遷移マップは話題距離を直感的に示し、対話状態グラフは対話のトピック距離の遷移構造を定量的に示す。

図 1 に本フレームワークのシステム構成図を示す。本フレームワークは対話テキストを入力とし、話題遷移マップと対話状態グラフを分析することで、対話の流れをいくつかの基本パターンに分類する。基本パターンの組み合わせにより、実際の対話の話題の収束、追従、膠着、停滞などの状態を判別できるか検証し、考察する。

なお、本稿で扱う対話は、チャット等のテキストデータを用いる。

4. 対話可視化フレームワークの構築

対話可視化フレームワークの具体的な処理の流れとして、まず対話から 1 つずつ発話を取得する。次に発話テキストから名詞を抽出し、単語埋め込みモデルで名詞を単語ベクトルに変換する。 ここで、単語の意味的な近さをノード間の物理的な近さとして反映させるため、あらかじめ約 2.6 万語の単語ベクトルとその 2 次元座標を格納した「背景知識マップ」を構築しておく。そして、先に変換した単語ベクトルをこの背景知識マップと比較し、意味的に近い単語群の周辺に配置されるよう 2 次元ベクトルを求める。この結果をノードとして話題遷移マップにプロットする。同時に、対話状態グラフを描画するため、発話トピック距離と対話トピック距離を計算する。発話トピック距離は、各話者がプロットした最新のノードとその直前のノード間の距離（例：赤色の New ノードと Previous ノードの距離）である。対話トピック距離は、2 者の最新のノード間距離（例：赤色の New ノードと青色の New ノードの距離）である。これらの距離を縦軸、時間を横軸とした折れ線グラフを描画する。

4.1. 名詞抽出・単語埋め込みモジュール

対話の中心的な内容と主題を捉えるため、本フレームワークは名詞に着目する。例えば、「昨日、駅で友達と映画について話したよ」という文において「昨日、駅、友達、映画」という名詞は、対話の概要を把握する上で重要な手がかりとなり得る。そのため、対話中に登場する名詞のみをノードとしてプロットする。また、名詞の中でも、一般名詞、固有名詞、サ変接続名詞に注目し、ノイズとなる可能性のある代名詞などを除外した。

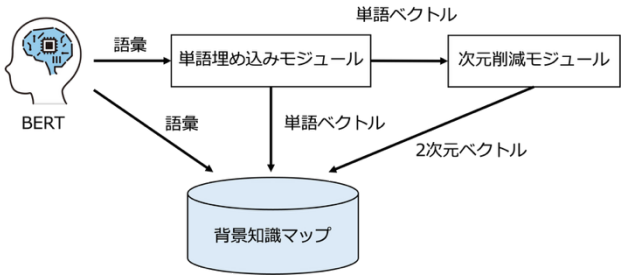


図 2: 背景知識マップの構築手順

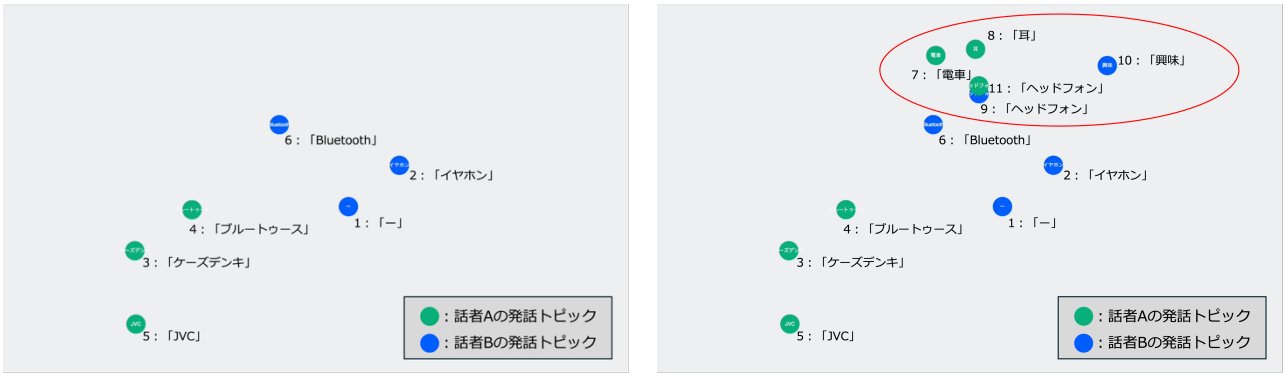
本研究では、発話テキストから名詞を抽出するために形態素解析器である MeCab を用いた。また、対話文という口語体のテキスト中には、新語や固有名詞が高い確率で登場すると考えたため、辞書に新語や固有名詞に強い NEologD を適用した。辞書に NEologD を適用した場合とデフォルト辞書を用いた場合の解析結果の比較を表 1 に示す。リモートワークやボカロ P、米津玄師など、NEologD を用いた場合の方が現代のテキストに含まれる固有名詞を取得できる。

対話可視化フレームワークは、発話テキストから抽出した名詞を、背景知識マップを用いることで意味的に類似している名詞の近くにプロットする。このとき、名詞同士が意味的に類似しているかどうか判別するために、単語ベクトルを用いる。単語ベクトル間のコサイン類似度を計算することで、単語同士の意味的類似度を判別する。

本研究では、名詞を単語ベクトルに変換するために、東北大学が公開している日本語事前学習済み BERT モデルを用いた。そして、モデルに一般常識的な語と語の関係の知識を持たせるため、Wikipedia の記事データで追加学習を行った。追加学習では、Masked Language Model (MLM)タスクを採用した。これにより、モデルが学習していない未知語に対しても意味を推論することが可能になる。

4.2. 背景知識マップ

発話トピックを 2 次元空間にプロットする際、単語間の意味的な近さをノード間の物理的な近さとして反映させることが、話題が近いことを示す上で重要である。この意味的な位置関係に基づいた安定的な配置を実現するために、本フレームワークでは「背景知識マップ」を導入する。



「時刻 1～6」　　**「時刻 1～6」＋「時刻 7～11」**

図 3: 話題遷移マップのプロットの様子

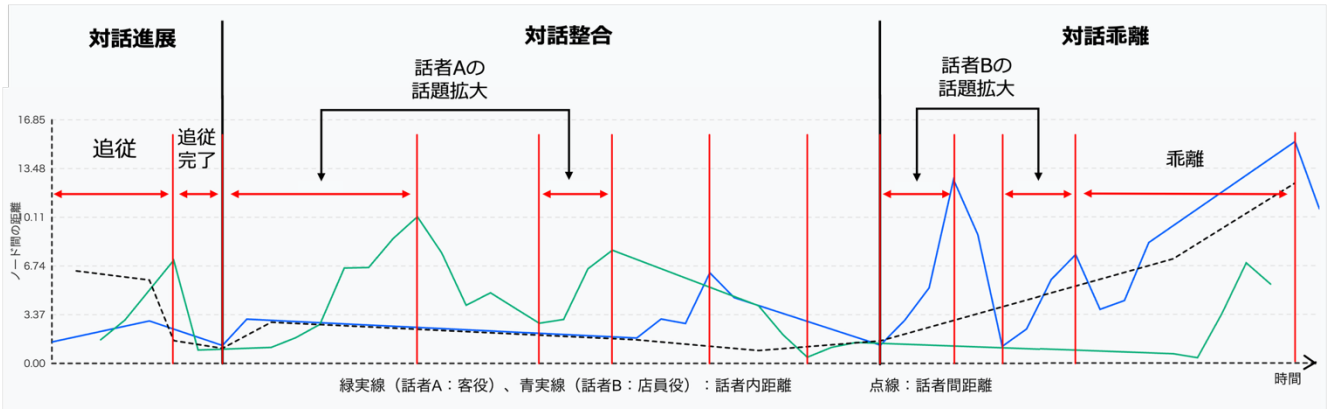


図 4: ヘッドホン購入対話における対話状態グラフ

これは、あらかじめ約 2.6 万語の単語と、それぞれの意味を表現する高次元の単語ベクトル、そして可視化のための 2 次元座標を紐づけて格納したものである。新しいノードを追加する際は、このマップを参照し、意味的に近い単語群の周辺に配置することで、前述したような自然な位置関係を実現する。

本研究では、辞書にない未知語や造語に対しても安定した意味的類推を行うため、この背景知識マップを構築するために 4.1 節で追加学習を行った BERT モデルの語彙を利用する。この語彙は一般的な単語だけでなく、より細かな意味を持つサブワードも網羅しており、サブワードを手がかりに意味的な類推が働くことで、安定した位置関係の構築が可能となる。

ベクトルの位置関係を直感的に理解するためには、2 次元または 3 次元のベクトルを用いる必要がある。しかし、本研究で用いる単語ベクトルは 768 次元であるため、次元削減を用いて次元数を 2 次元にまで落とし込む必要がある。本研究では、データの局所的な位置関係を捉えるために t-SNE の手法を用いた。

背景知識マップの構築の流れを図 2 に示した。まず、学習した BERT モデルが単語埋め込みの際に参照している語彙を取得する。次に、これらの単語を BERT モデルで単語ベクトルに変換する。そして、t-SNE で次元削減を行うことで、2 次元ベクトルまで次元を落とし込む。語彙とそれに対応する単語ベクトル、2 次元ベトルの組を、背景知識マップとして保存する。

4.3. 話題遷移マップ

話題遷移マップは、発話トピック遷移の構造を直感的に理解することを目的とした 2 次元空間である。この空間では、意味的に近い名詞は近い位置にプロットされるため、それぞれの話者が発話した名詞ノードが近い位置にプロットされていれば、話している話題が近いことがわかる。逆に、ノードが徐々に遠ざかっていけば、話題が離れていっていると把握できる。このため、ノード（名詞）の位置変化を見ることで、対話の展開や噛み合い方を把握することが可能となる。

図 3 に、対話パターンの一例である「追従」の様子を示した。ノードの色は話者ごとに変えている（例：話者 A が緑色、話者 B が青色）。 図 3 の発話順 1 番目から 6 番目までは、それぞれの話者の発話

トピックは離れた位置にプロットされている。この区間では、話者 A（緑）は「イヤホン」「Bluetooth」、話者 B（青）は「ケーズデンキ」「JVC」などについて発話しており、互いの話題には距離がある。 しかし、7 番目から 10 番目にかけて、話者 A（緑色ノード）の発話トピックが、話者 B（青色ノード）の発話トピックに近づいていることが確認できる。この区間では、話者 B が「ヘッドフォン」について言及したのに対し、話者 A も「電車」「耳」といった単語を経て「ヘッドフォン」について発話している。この視覚的な位置変化から、話者 A が話者 B の発話トピックに対して話題を合わせる「追従」が発生していると分析できる。 このように、ノードの位置変化を見るだけで、対話がどのように展開しているのかを直感的に捉えることが可能になる。

この話題遷移マップを構築する上で、リアルタイムで発話される名詞を、既存のノードとの意味的な位置関係を壊さずにプロットする必要がある。

ノードを追加するたびに t-SNE を用いて全ノードの次元削減を再計算する方法も考えられる。しかし、t-SNE は確率的に次元削減を行うため、新たな計算結果が元の 2 次元ベクトルと必ずしも一致するとは限らず、元の位置関係が崩れてしまう課題があった。

この課題を解決し、元の位置関係を保持したまま新たなノードを安定して配置するため、本研究では k-近傍法を用いる。k-近傍法は、新たに追加するノードの単語ベクトルと意味的に類似度の高い上位k個の近傍点（背景知識マップ上の点）に基づいて 2 次元座標を決定する。これにより、意味的な関係性を捉えたまま（＝意味的に近いノードの近くに）新たな単語を配置することが可能になる。

4.4. 対話状態グラフ

対話状態グラフは、各話者の発話トピックが意味的に類似しているか、乖離しているかの時間的変化を定量的に判断できる折れ線グラフである。

本研究において、ノードは単語ベクトルから作成しているため、ノード間のユークリッド距離は単語の意味的類似度を表す。距離が近ければ (0 付近であれば）意味的に近く、距離が遠ければ意味的に乖離していると見なせる。 この距離の変化を分析することで、追従や話題拡大、乖離などの状態を判別することが可能になる。

図 4 にヘッドホン購入対話における対話状態グラフを示した。グラフの横軸は時間、縦軸はノード間の距離である。実線はそれぞれ話者 A、話者 B の発話トピック距離の変化、点線は対話トピック距離の変化を示している。発話トピック距離と対話トピック距離から対話の状態を以下のパターンとして分類できる。

- **追従**
  本研究での「追従」とは、片方の話者の発話トピックに対して、もう片方の話者が自分の発話トピックを合わせている状態である。
  これは図 4 の「追従完了」区間のように対話トピック距離が小さい区間において、片方の発話トピック距離が常に小さいまま推移し、もう片方が大きい値から小さい値へと変化する場合や両者の発話トピック距離が同じように増減する場合に確認できる。
- **話題拡大**
  本研究での「話題拡大」とは、話者が対話中の自分の発話トピックを変更している状態である。
  これは図 4 の「話者 A の話題拡大」「話者 B の話題拡大」区間のようにグラフ上において、話者の発話トピック距離が大きくなる変化として確認できる。
- **乖離**
  本研究での「乖離」とは、両方の話者の発話トピックが離れている状態である。
  これは図 4 の「乖離」区間のように対話トピック距離が大きい区間において、両者の発話トピック距離が同じように増減する場合や、片方の発話トピック距離が常に小さいまま推移し、もう片方の距離が大きく変化する場合に確認できる。

このように、発話トピック距離と対話トピック距離から対話の状態を分類することで、話題遷移マップによる直感的な理解を裏付け、定量的に対話の状態を分析することが可能になる。

5. ケーススタディ

5.1. 実験概要

実際の対話における対話状態グラフがどのパターンの組み合わせで構成されているのか分析するため、複数名の被験者にヘッドホンの新規購入をテーマとした対話を行ってもらった。

実験の設定として、対話ではヘッドホンを新規購入する客役、ヘッドホンを紹介する店員役とした。また、店員役は商品を薦めるため、インターネットから情報を参照してもよいこととした。被験者は制限時間 10 分の間に、最終的にヘッドホンを購入するかどうか決める。対話は LINE のメッセージ機能で行った。

5.2. 結果と考察

図 4 の対話状態グラフで割り当てられている状態が、実際の対話テキストの内容と整合しているかを検証する。

なお、本検証において「追従」と「追従完了」は話題が収束していく一連のプロセスであるため、「追従」としてまとめて扱う。また、「話者 A の話題拡大」と「話者 B の話題拡大」は、話者が異なるのみでグラフ上の挙動および対話における相互作用の構造は同一であるため、「話題拡大」としてまとめて扱う。

1. 追従および追従完了

実際の対話テキストでは、挨拶の後、店員が「今イヤホンは何を使っているか」と問いかけ、客が「ケーズデンキ」「JVC」と具体的な製品名を挙げて回答している。これは、店員の問いかけに対し客が具体的な情報を提示することで、双方が「客の現在の利用状況」という共通の話題に焦点を合わせた過程を示している。

その後、客の回答を受けて店員が「Bluetooth 使いやすくていいよね」と肯定的な感想を述べており、話題が完全に一致した状態で会話が進んでいる。これらの流れは、グラフにおける距離の縮小および安定した推移と整合しており、判定は妥当である。

2. 話題拡大

実際の対話テキストを確認すると、客が「紛失への不安」「過去の耳の痛みの記憶」「音質への興味」など、自身の悩みや要望に関連する新しい要素を次々と発話している。これに対し、店員（話者 B）は客の話を受け止める側に回っているため、距離の変化は小さい。

このように、話者が能動的に話題を広げている状況が、グラフ上での距離の拡大として正しく反映されており、判定は妥当である。

3. 乖離

実際の対話テキストでは、「音質」「色」「Amazon」「タイムセール」「購入決定」と、意思決定に向けて短期間に多岐にわたるトピックが交換されている。

文脈としては購入の合意形成に至っており対話は成立しているが、本システムは名詞の意味的類似度に基づいているため、「色」と「EC サイト」のように意味的に距離の遠い単語が連続すると「乖離」として検出される。これは、意思決定の最終段階において、必要な情報を網羅するために多様なトピックが飛び交う際に見られる特徴的なパターンであると解釈できる。

6. まとめ

本研究では、意思決定対話における認識のずれや話題の遷移を客観的に捉え、円滑な対話を支援するため、トピック間距離に基づいた対話可視化フレームワークを提案した。

本手法は、発話トピックを単語ベクトルとして扱い、話題遷移マップによる直感的な可視化と、対話状態グラフによる定量的な分析を可能にするものである。実際の対話データを用いたケーススタディの結果、グラフから読み取れる「追従」「話題拡大」「乖離」といった状態が、実際の対話内容と整合していることを確認し、本フレームワークの有効性を示唆した。

今後の課題として、より詳細な対話状態の判別精度の検証が挙げられる。具体的には、話題の収束、発散、追従、減退、膠着、停滞などの状態をアノテーションした対話データセットを用い、本システムがこれらの状態を正しく判別できるか検証する実験を行う。また、対話状態グラフの解析において、話者内距離の傾きを正、零、負のいずれに分類するかを判断するための明確な閾値を、予備実験を通じて決定していく予定である。

参考文献

[1] 角康之, 小川竜太, 堀浩一, 大須賀節雄, 間瀬健二: 思考空間の可視化によるコミュニケーション支援手法 CSS, 電子情報通信学会論文誌 A, Vol. 79, No. 2, pp. 251-260, (1995)

[2] 角康之: 議論の意味構造の可視化, 可視化情報, Vol. 19, No. 72, (1999)

[3] Ryotaro Okada, Takafumi Nakanishi, Yuichi Tanaka, Yutaka Ogasawara, Kazuhiro Ohashi: A Time Series Structure Analysis Method of a Meeting Using Text Data and a Visualization Method of State Transitions, New Generation Computing, Vol. 37, pp. 113-137, (2019)

[4] 柳楽浩平, 水本武志: 話し合いの振り返りのためのオンラインや対面の会話の定量化と可視化, The 37th Annual Conference of the Japanese Society for Artificial Intelligence, (2023)

多角的視点を持つマルチエージェントシステムによる要件定義レビュー

Multiple Agents with Different Roles for Reviewing Requirements Definition

井上 祐寛 [1]　　松永 嵩 [1]　　綾塚 祐二 [1]

Takuhiro Inoue[1],　　Takashi Matsunaga[1],　　Yuji Ayatsuka[1]

[1] 株式会社クレスコ 技術研究所
[1] Technology Laboratory, CRESCO, LTD

We propose using multiple agents to review the requirements definition to improve quality. Each agent plays a different role, such as project manager, business analyst, or system architect, and performs cross-checking from different perspectives to reduce oversights, false detections, and contradictions. As a first step, we constructed a review agent that works as a project manager using an large language model (LLM) and evaluated its output.

1. はじめに

ソフトウェア開発プロジェクトにおいて、要件定義は全体の成否を左右する最も重要な初期工程の一つである。この段階での不備、特に要求の抜け漏れや曖昧な記述は、後続の設計・開発工程で大規模な手戻りを引き起こし、開発コストの増大と納期の遅延に直結する。このリスクを低減し要件定義の品質を向上するため、従来より多様な視点を持つステークホルダー（たとえば、ビジネス部門、開発部門、品質保証部門など）によるレビューが不可欠とされてきた。

しかし、多様な視点を確保するためには、専門知識を持つ多くの担当者のリソースを確保する必要があるが、人数が増えるほどスケジュール調整は困難となる。また、担当者の専門領域に起因する視点の偏りといった属人的な課題も常に存在し、網羅的な品質担保の障壁となっている。

本研究では、これらの課題を解決するため、Large Language Model (LLM) を活用し、個々の LLM エージェントに対して「プロジェクトマネージャ (PM)」「ビジネスアナリスト」「システムアーキテクト」といった異なる専門的役割（ロール）を割り当てることで、多角的な視点を持つマルチエージェントによるレビューシステムを提案する。

LLM の役割演技（ロールプレイング）能力を活用し、各エージェントがそれぞれの専門的視点から要件定義書を並列でレビューする。これにより、単一の視点では見逃されがちな課題（例：機能要件の矛盾、非機能要件の欠落、ユーザビリティやテスト容易性の課題）を網羅的に洗い出し、また、エージェント間の相互照合により誤検出や矛盾の低減を目指す。

本稿は、その最終目標に向けた第一段階の研究報告である。本稿では、まずシステムの基盤として「プロジェクトマネージャ(PM)」の役割を付与した単一エージェントが動作する実験環境を構築する。そして、エージェントによるレビューの実行手順と、その生成物の品質を定量的に評価するための「レビュー手順と評価の枠組み」を整備する。この枠組みの有効性を検証するために実施した、仕様書を用いた実験結果についても併せて報告する。

2. 関連研究

本研究は、「LLM の役割演技」、「LLM マルチエージェントシステム（MAS）」、そして「LLM による多視点評価」という、近年急速に進展している三つの研究領域を融合し、ソフトウェア工学の「要件定義レビュー」というドメインに適用する試みである。

2.1. LLM の役割演技（LLM Role-Playing）

LLM に特定の役割（ロール）を割り当てる「ロールプレイング」は、LLM の能力を引き出す有効な手法として確立されている。Tseng らのサーベイ[1]では、LLM が割り当てられた役割に基づき、その環境や文脈に適応した応答を生成する能力（LLM Role-

Playing）が示されている。本研究ではこのロールプレイング技術を応用し、ソフトウェア開発の各ステークホルダーとしての専門的視点を持つエージェントを構築する基盤とする。

2.2. LLM マルチエージェントシステム (MAS)

単一の LLM では解決困難な複雑なタスクに対し、複数の LLM エージェントが協調して問題解決にあたるマルチエージェントシステム（MAS）の研究が進展している。Guo らのサーベイ[2]によれば、MAS は各エージェントに特化した役割と知識を与えることで、集合的な知性を活用するアプローチである。

特にソフトウェア開発領域では、MetaGPT[1] や AutoGen[2] といったフレームワークが提案されている。これらは、エージェントに「プロダクトマネージャ」、「プログラマー」、「テスター」といった役割を割り当て、コーディングやテストといった開発プロセス自体を自動化する試みである。これらの先行研究が主に「開発プロセスの自動化」に焦点を当てているのに対し、本研究は MAS のアーキテクチャを「要件定義レビュー」という特定の品質保証タスクに応用する点に特徴がある。

2.3. LLM による多視点評価

LLM に多様なペルソナを与え、評価タスクに適用する研究も進んでいる。南雲らの研究[3]では、ビジネスアイディアの評価において、多様な属性を持つペルソナが、それぞれ独自の評価基準を用いて評価を行うデルファイ法が提案されている。さらに、ファシリテーター役の LLM が各ペルソナの評価を集約・要約するプロセスも示されている。

本研究の最終目標は、この多視点評価の手法を「要件定義レビュー」というドメインに特化させることである。南雲らの研究が「属性（例：年代、性別）」に基づくペルソナを用いたのに対し、本研究では「職務的役割（PM、アーキテクト等）」に基づく専門的視点を用いる点で独自性がある。

3. 提案手法：レビュー手順と評価の枠組み

本稿では、PM 役割の単一エージェントによるレビュー環境と、その生成物の品質を定量的に評価する枠組みを設計・構築した。この枠組みは、図 1 に示す通り、二段階の LLM エージェント・パイプラインで構成される。この二段階構造の採用は、LLM の「生成」と「品質検証」の役割を分離し、生成されたレビューを定義した点数付けに従って評価可能とするためである。これにより、レビュー指摘の品質を定量的に判断できることを確認し、評価プロセスの安定化を図る。

3.1. 第一段階：レビュー生成エージェント (PM Agent)

第一段階は、仕様書をレビューし、指摘事項を生成するエージェントである。

・役割

エージェントには、LLM のロールプレイング能力を活用し、明確な役割を付与した。具体的には、「IT 企業入社 30 年目（開発 20 年、プロジェクトマネージャ 10 年）の経験豊富なプロジェクトマネージャ」と定義した 。

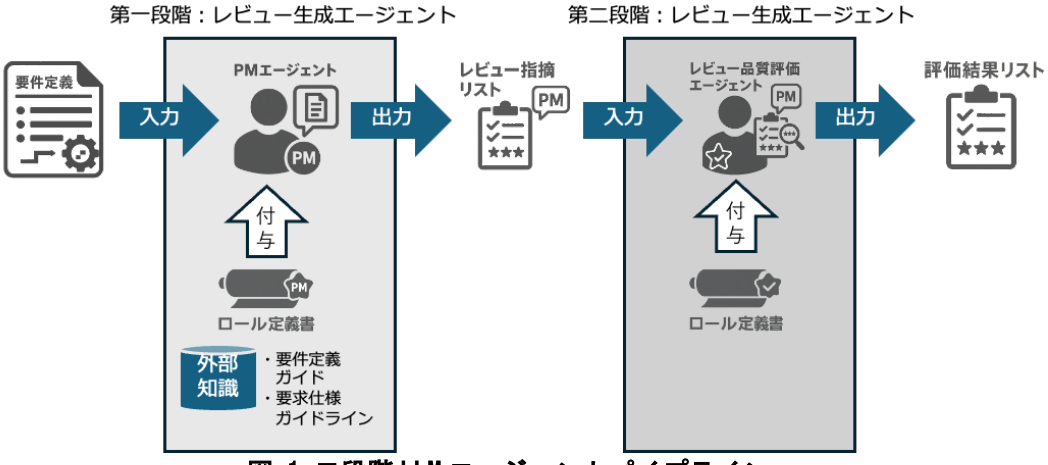


図 1. 二段階 LLM エージェントパイプライン

1 MetaGPT https://github.com/FoundationAgents/MetaGPT

2 AutoGen https://microsoft.github.io/autogen/0.2/

・タスク
役割定義に基づく専門的視点から、入力された要件定義書（本研究では「話題沸騰ポット要求仕様書（GOMA-1015 型）第 3 版[3]」）をレビューする。その際、出力形式は「修正番号」、「指摘箇所」、「指摘した記述」、「指摘の理由」、「指摘箇所の添削例」に統一することとした 。

・基盤モデルと知識
エージェントの基盤モデルには、OpenAI の ChatGPT モデル群の一つであり、複雑な推論に強いとされる`o3`モデルを採用した。また、エージェントの専門性を担保し回答の質を向上させるため、外部知識として IPA（情報処理推進機構）の「ユーザのための要件定義ガイド」および JUAS（日本情報システム・ユーザー協会）の「要求仕様ガイドライン」を付与した。

・レビュー抽出手法
レビューの生成方法として、一度に多数（例：100 件）を要求する手法と、対話的に少数ずつ深掘りする手法を比較した。その結果、一度に多数を要求すると、修正理由や修正案が簡素化される傾向が観測された。そのため本研究では、「他にありますか？」「もっとありませんか？」と対話的に深掘りする手法を採用し、エージェントが持つ指摘事項を網羅的に（本実験では 40 件）抽出した

3.2. 第二段階：レビュー品質評価エージェント（Scoring Agent）

第一段階で生成されたレビューの「品質（＝指摘の妥当性や重要度）」を定量的に採点するエージェントである。この採点は、指摘の重要度を明確化し、レビュー後の仕様書修正の優先順位付けに必要な情報を提供するために行う。

・役割
採点エージェントにも、レビュー生成エージェントと同様のペルソナ（経験豊富な PM）を付与した

・タスク
第一段階で生成された個々のレビューコメントに対し、要件定義書と照合した上で「辛口」で 0〜5 点の整数採点を行うよう指示した。

・採点基準

評価の客観性を担保するため、厳格な採点基準を定義した（表 1）

表 1.採点基準

点数	判断基準（概要）	詳細な意味合い
5	致命的	基本仕様・安全要件の欠落や重大法令違反。製品化停止レベル、最優先で是正。
4	重大	法規・規格違反リスク、事故・故障・市場クレームに直結。設計手戻り大。
3	中度（要修正）	機能・保守・安全・規格適合に実質影響。放置不可。
2	軽度	品質・操作性へ間接影響。回避策・部分記載あり。
1	ごく軽微	誤字・体裁、機能・安全に影響せず即日修正可。
0	指摘無効	すでに仕様書に記載／指摘不要／誤認。

・評価の安定化
LLM による採点は、同一の入力に対しても「ばらつき」を生じる可能性がある。この問題を軽減し評価の信頼性を高めるため、本研究では同一のレビューコメントに対して採点エージェントを 5 回独立して実行し、その平均点を最終的な品質スコアとして採用した。

4. 実験：評価枠組みの適用と検証

4.1. 実験目的

本稿で構築した「レビュー手順と評価の枠組み」（3 章）が、要件定義レビューにおいて有効に機能するかを実証する。具体的には、第一段階の PM エージェントが実用的な指摘を生成できること、および第二段階の採点エージェントがその指摘品質を妥当に定量化できることを検証する。

4.2. 実験設定

実験対象として、ハードウェアの仕様書（「話題沸騰ポットの仕様書 V3」）を用いた 。この仕様書は、実際の製品開発で用いられるレベルの詳細度を持ちつつ、レビューの観点からは (意図的な)曖昧さや欠落箇所が含まれている 。

3 組込みソフトウェア管理者・技術者育成研究会(SESSAME)　https://www.sessame.jp/
電子ポットを題材にした組込みシステム分析・設計のための要求仕様書

実験手順として 3 章で定義した二段階評価パイプラインを適用した。

（第一段階）PM エージェントが対象仕様書をレビューし、対話的な深掘りによって 40 件の指摘事項を生成した。

（第二段階）採点エージェントが、生成された 40 件の指摘事項をそれぞれ 5 回ずつ独立して採点し、平均点を算出した 。

4.3. 実験結果

本枠組みによる評価結果は、ハードウェア仕様書のレビューとして極めて妥当なものであった。採点エージェントは、プロジェクトリスクに直結する「致命的な欠陥」の指摘には一貫して高得点（平均 4.8 点以上）を付与し、一方で「改善提案」レベルの指摘には低得点を付与する、明確な傾向を示した。

(1) 高スコアの指摘

表 2 はスコア（平均 4.8 点以上）を獲得した重要指摘の例である。安全性、基本機能などに関する致命的な欠落が含まれ、PM エージェントが、その役割通り、プロジェクトの根本的なリスク(安全、基本機能不全)を最優先で特定できていることを示す。

(2) 低スコアの指摘

表 3 はスコア（平均 3.0 点未満）となった指摘の例である。主に詳細なユーザビリティ（使い勝手）に関する仕様の改善提案が含まれる。これらは「あれば望ましい」ものであり、PM の視点からは（表 2 の致命的欠陥と比較して）優先度が低いと判断されている。採点エージェントがこの重要度の差を正しくスコアに反映できていることを示す。

5. 考察

本研究で構築した二段階の評価枠組みを適用した結果、提案手法の有効性と、今後のマルチエージェント化に向けた明確な課題が示された。

5.1. PM エージェントの実用性

実験結果（表 2）が示す通り、構築した PM エージェントは、安全性（感電、破裂リスク）、基本仕様（電源、貯水容量、防水等級）といった、プロジェクトの致命的な欠陥を優先的に特定する能力を有していることが実証された。これは、PM という役割定義と、外部知識の付与が有効に機能したことを示唆し

表 2：高スコア（平均 4.8 点以上）を獲得した重要指摘の例

修正理由	修正案	平均点
アース端子/漏電保護など感電対策記載無し	三極プラグ＋漏電ブレーカ内蔵を必須、安全規格(IEC 60335)準拠を追記	5
蒸気排気路/過圧逃がし構造記載無しで破裂リスク	0.05 MPa で開く弁構造を追加し試験プロトコルを図示	5
最大/最小貯水容量(L)が仕様に見当たらない	公称 1.5L、最小加熱水量 0.3L など容量レンジを追加	4.8
接水樹脂部の食品衛生法・BPA フリー等の材質要件が欠落	口金・蓋パッキンは厚生労働省告示第 370 号適合材を指定	4.8
防水・防滴(IP)等級が未規定で台所利用時に安全性不足	最低 IPX1、推奨 IPX4 の筐体設計と試験条件(IEC60529)を明記	4.8
AC100V/50-60Hz など電源電圧・周波数・許容変動の記載無し	定格 100V±10% 50/60Hz、突入電流・待機電力上限を明記	4.8

表 3：低スコア（平均 3.0 点未満）となった指摘の例

修正理由	修正案	平均点
タイマは 1 分刻みのみで 30 秒以下の短時間設定が不可能	秒単位の加算/長押し高速算モードを追加	2.6
タンは 1 分加算のみで長押連続入力・チャタ対策時間未定義	押し 0.8s 以上で 10 分刻み算、離した瞬間に停止と明	2.8
イマ上限「最大 1 時間」の拠不明	ースケースに基づき上限値妥当性を説明 or 可変設定	2.4
ザー音量固定で夜間利用に慮不足	ニューで 60dB⇔70dB 切、または消音＋LED 通知追加	2.8

ており、単一のエージェントであっても実用的なレビューエージェントとして機能しうることを示している。

5.2. LLMによる品質評価枠組みの有効性

本研究の評価枠組みの中核である「レビュー品質評価エージェント（辛口採点エージェント)」は、その妥当性を実証した。採点エージェントは、PMエージェントが生成した指摘事項に対し、致命的な欠陥（表2）には一貫して高い平均点（4.8点以上）を付与し、ユーザビリティ等の改善提案（表3）には低い平均点（3.0点未満）を付与した。これは、LLMが定義された採点基準に基づき、指摘事項の品質（重要度や妥当性）を定量的に評価可能であることを示している。また、採点を5回繰り返して平均する手法は、LLM固有の「ばらつき」を抑え、評価の信頼性を担保する有効な枠組みであると結論付けられる。

5.3. 単一エージェントの限界とマルチエージェントシステムの必要性

本実験の考察で最も重要な点は、単一エージェントの限界が明確になったことである。PMエージェントは、その役割通り「プロジェクトの重大リスク」を優先する一方で、表2に示すような「詳細なユーザビリティ（タイマー刻み、ブザー音量)」に関する指摘の優先度は低く、見落とされる可能性が考えられる。

これは、単一の視点ではレビューが偏るという従来（人間によるレビュー）の課題を裏付けるものであり、本研究の最終目標であるマルチエージェント・レビューシステムの必要性を強く示唆している。PMエージェントが見落とす可能性のある「ユーザビリティ」の観点は、まさしく「UXデザイナー」や「ユーザー」といった異なる役割を持つエージェントが補完すべき領域である。

6. 結論と今後の展望

6.1. 結論

本研究では、LLMによる多角的な要件定義レビューシステムの構築に向けた第一段階として、「プロジェクトマネージャ（PM)」役割の単一エージェントによるレビュー手順と、その品質をLLMで定量的に評価する評価の枠組みを構築・整備した。

提案する枠組みは、以下の二段階パイプラインで構成される。

・第一段階（レビュー生成）
明確な役割（PM）と外部知識を付与された「PMエージェント」が、対話的な深掘りを通じて仕様書の指摘事項を生成する。

・第二段階（品質評価）
同様の役割を持つ「辛口採点エージェント」が、厳格な基準に基づき、複数回（5回）の採点平均によってレビュー品質を安定的に定量化する。

実験により、本枠組みが有効に機能することを実証した。PMエージェントは安全性や基本仕様に関する重大な欠陥を優先的に特定し(平均スコア4.8点以上)、採点エージェントがその品質を妥当に評価できることを確認した。

同時に、PMエージェントの指摘がその役割に偏ることで、ユーザビリティ等の観点が見落とされる可能性という単一エージェントの限界も明確になった。この結果は、本研究の最終目標であるマルチエージェントシステムの必要性を強く裏付けるものである。

6.2. 今後の展望

今後の研究では、本稿で明らかになった単一エージェントの限界を克服すべく、提案するマルチエージェント・レビューシステムの構築を進める。

まず、PMに加え、ビジネスアナリスト、システムアーキテクトなどの多様な役割を持つエージェントの実装に取り組む。そして、各エージェントが独立したレビュー結果を共有し、矛盾を解消するための協調メカニズムの構築を主要なテーマとする。

また、システムの信頼性向上と機能拡張に向けた研究にも並行して取り組む。これには、自己肯定バイアスやハルシネーションの抑制効果を定量的に測定する信頼性の客観的検証が含まれる。将来的には、より高度な論理推論や判断を可能にするための新たな技術的アプローチも継続して検討していく。

参考文献

[1] Yu-Min Tseng, Yu-Chao Huang, Teng-Yun Hsiao, Wei-Lin Chen, Chao-Wei Huang, Yu Meng, and Yun-Nung Chen. "Two Tales of Persona in LLMs: A Survey of Role-Playing and Personalization." [cite_start]*arXiv preprint arXiv:2406.01171v3 [cs.CL]*, 2024. [cite: 3109, 3110-3116]

[2] Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V. Chawla, Olaf Wiest, and Xiangliang Zhang. "Large Language Model based Multi-Agents: A Survey of Progress and Challenges." [cite_start]*arXiv preprint arXiv:2402.01680v2 [cs.CL]*, 2024. [cite: 2006, 2007-2010]

[3] 南雲陸, 佐々木. "LLM を活用したペルソナベースのデルファイ法による多視点アイディア評価." [cite_start]*第 39 回 人工知能学会全国大会 (The 39th Annual Conference of the Japanese Society for Artificial Intelligence)*, 2025. [cite: 1860, 1861-1864, 1867]

[4] Zefang Zong, Jingwei Wang, Yunke Zhang, et al. "Towards Large Reasoning Models: A Survey of Reinforced Reasoning with Large Language Models." [cite_start]*arXiv preprint arXiv:2501.09686v3 [cs.AI]*, 2025. [cite: 2131, 2132-2137]

複数領域に対するキャプション生成を用いた目の不自由なユーザ向けの画像理解支援

A System for Image Understanding Support for Visually Impaired Users Using Multi-Region Caption Generation

XU Yiling[1]　SHAN Junjie[1]　安尾 萌[2]　西原 陽子[1]
Yiling Xu[1], Junjie Shan[1], Megumi Yasuo[2], and Yoko Nishihara[1]

[1] 立命館大学　情報理工学部
[1] College of Information Science and Engineering, Ritsumeikan University
[2] 立命館グローバル・イノベーション研究機構
[2] Ritsumeikan Global Innovation Research Organization

Abstract: This study proposes an interactive support system designed to assist visually impaired users in achieving a deeper understanding of complex images. Conventional methods that generate a single holistic caption often fail to convey complex compositions. To address this issue, we propose a sub-image captioning approach. We implemented and evaluated two methods: the Simple Sub-image Captioning Method, and the Max Cover Dense Captioning Method. Experimental results demonstrated that the latter method achieved higher objective scores in eight out of 10 image categories, particularly for images with complex scenes and small but critical elements, and a similar trend was observed in the subjective evaluations. Conversely, for images where a single item occupies most of the frame, a holistic description sometimes proved more effective, indicating that the optimal strategy is content-dependent.

1　はじめに

視覚に不自由があることは単なる公衆衛生上の深刻な課題であるだけでなく，社会的な公正性と人類の福祉に関わる重要な議題でもある [1, 2]．そのため，目の不自由なユーザ向けユーザインタフェース（UI）の開発は，益々喫緊の課題となっている [3]．

生成 AI などの近年の技術革新に支えられ，これらのツールは彼らの情報化社会への参加に新たな機会を提供している．その中核となるアプローチは，画像内の視覚情報をテキスト記述に変換し，それを音声合成を介して読み上げることで，ユーザが周囲の世界や任意の画像を「聞く」ことを可能にするものである [4, 5]．

既存技術は単一文章の生成において大きな成功を収めているが [6, 7]，画像全体に対して包括的な説明文を生成するという主流のアプローチには，情報の伝達能力において深刻な限界が存在する．

実際のユーザを対象とした調査では，既存の AI ツールが生成する記述は「詳細さが不十分」であり [8]，物体の空間的な位置関係や色といった具体的な視覚情報，あるいは人物間の関係性のような文脈情報が欠落しているという不満が広く報告されている [9]．これは，従来の手法が複雑な画像情報を単一の文章に圧縮する過程で，ユーザが明確なメンタルイメージを構築するために不可欠な詳細情報を失ってしまうためである．さらに，情報は一方的に提示されるため，ユーザが個人的に興味のある領域を対話的に探索し，詳細を得ることもできない．したがって，多領域的かつ構造化された画像情報を対話的に提示する手法を開発することが，これらの問題を解決する鍵であり，本研究の主な焦点である．

本研究では，単一の説明文による限界を克服する画像理解支援システムを提案する．具体的には，画像を複数のサブ領域に分割し，各領域に対して独立した説明文を生成する，新たな「サブ画像記述」手法を導入する．この手法は，目の不自由なユーザがより豊かで詳細なメンタルイメージを構築し，脳内でより現実に近い視覚シーンを再構成できるよう支援することを目的としている．

2 関連研究

本章では，提案システムの研究に関連する既存研究を，(1) 目の不自由なユーザ向け支援システムの応用研究と，(2) 画像説明生成の基盤となるマルチモーダルモデルの技術研究，の 2 つの観点から概観する．それぞれの分野における既存研究の達成点と未解決の課題を明らかにすることで，本研究の位置付けを明確にする．

2.1 目の不自由なユーザ向け支援システムの既存研究

近年，目の不自由なユーザが画像を理解するために，どのような記述を提供すべきかについて活発な研究が行われている．Doore ら [10] は，特にアート作品の鑑賞という文脈において，ユーザの要求と AI モデルの性能を包括的に調査した．彼らはユーザ調査を通じて，短い一行の概要説明だけでは不十分であり，空間情報と主題情報の両方を含む詳細で多層的な記述が強く好まれることを明らかにした．

また，Fernando ら [8] は，日常的に利用される画像認識ツール（IRT）に関する広範なレビューとユーザ評価を行った．その調査結果によれば，ユーザが既存ツールに対して抱く主な不満点は「記述が不十分」であることであり，将来のツールに最も望む改善点として「より詳細な情報」の提供が挙げられている．

このように，芸術鑑賞という特殊な文脈と，日常利用という一般的な文脈の双方において，既存の単一文章による画像説明がユーザの「深い理解」へのニーズを満たせていないという共通の課題が確認された．この事実は，本研究で提案するサブ画像キャプション生成のような，より構造化され詳細な情報を提供するための新しいアプローチの必要性を強く示唆している．

2.2 画像説明生成マルチモーダルモデルに関する既存研究

2.1 節で述べた先進的な応用を実現するには，深層学習に基づく画像説明生成分野の飛躍的な進歩が不可欠である．現在，この分野の進化は，OpenAI によって開発された CLIP（Contrastive Language-Image Pre-training）[11] に代表される，大規模視覚言語モデルの登場により新たな段階に入った．

CLIP がもたらした新しいパラダイムを応用した代表的な研究として，Mokady らが提案した ClipCap[12] が挙げられる．本研究では，この ClipCap の高い効率性と汎用性に着目し，提案システムのアーキテクチャとして採用する．

また，画像の局所的な理解を深める研究として，Johnson らによる「Dense Captioning」および DenseCap モデル [13] が挙げられる．これは従来の「1 画像に 1 説明」という枠組みを超え，画像内の多数の重要領域を自動特定し，各領域に個別の説明文を生成するものである．さらに Delloul ら [14] は，DenseCap を応用し，RGB-D カメラの深度情報を活用して物体間の位置関係（左右・前方など）を明確に記述する手法を提案した．しかし，この手法は特殊なハードウェアに依存するため，一般的な RGB 画像には適用できないという制約がある．

これに対し，本研究で提案するシステムは，画像の部分領域から生成された多数のテキスト記述を最適化し，ユーザへ提示するものである．提案システムは，(1) 深度カメラのような特殊なハードウェアを必要とせず，あらゆる RGB 画像に適用可能な汎用性を持つ点，そして (2) 単なるアルゴリズムの提案に留まらず，実利用を想定したユーザインターフェースの実装と評価を含んでいる点において，より実用的な貢献を目指すものである．

3 提案システム

2 章で述べた課題に対処するため，本研究では図 1 に示すように，複数の記述を用いて画像の理解を支援するシステムを提案する．本章では，この共通基盤の上に構築された，アプローチの異なる 2 つの具体的な手法について詳述する．

3.1 提案手法 1：部分領域記述法

1 つ目の提案手法は,「部分領域記述法」である．図 2 にその処理例を示す．本手法では，入力画像を 3×3 の均等なグリッドで一律に分割し，9 つのサブイメージ（領域）を生成する．続いて，これら 9 つのサブイメージをそれぞれ独立した画像として扱い，ClipCap モデルに入力することで，各領域に対応する個別の説明文を生成する．本手法の利点は，画像の全領域を網羅的にカバーできる点，および「左上」「中央」「右下」といったユーザにとって直感的かつ予測可能な構造で情報を提供できる点にある．

3.2 提案手法 2：重ね領域最大法

2 つ目の提案手法は，意味的文脈を考慮した「重ね領域最大法」である．図 3 にその処理例を示す．本手法の目的は，単純な部分領域記述法が持つ構造的な網羅性と明快さを維持しつつ，より意味内容の豊かな領域に対して説明を割り当てることにある．単純なグリッ

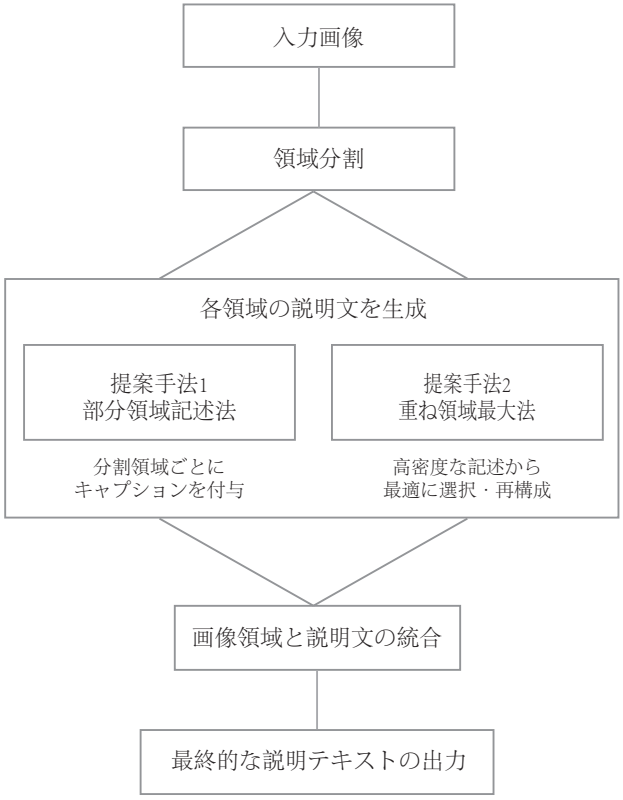


図 1: システムの概要

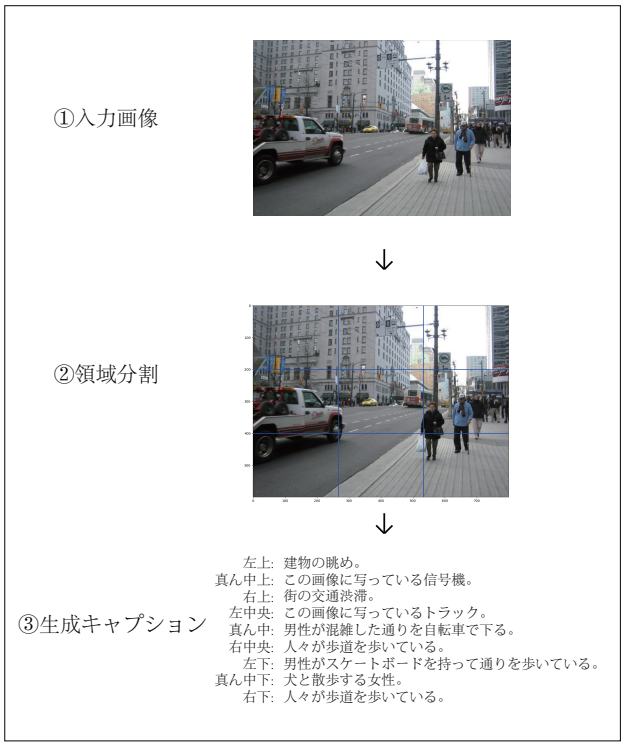


図 2: 提案手法 1 の実行例．画像は Visual Genome データセット [15] から引用．

ド分割では，物体が複数の領域に分断されたり，ある領域に意味のない背景のみが含まれたりするという課題がある．この問題を解決するため，本手法は以下の 2 段階のプロセスで構成される．すなわち，(1) 画像から詳細な領域記述候補を網羅的に生成する段階，および (2) 生成された候補の中からグリッド構造に合わせて最適な記述を選択する段階である．

第一段階では，Johnson らが提案した「Dense Captioning」のアプローチを適用する．まず，CNN を用いて画像から特徴マップを抽出する．次に，特徴マップ上の各位置を基準に様々なアスペクト比を持つバウンディングボックスを生成し，物体が存在し得る領域を提案する．モデル内の Localization Layer が各提案領域の信頼度を予測し，最終的に 1 枚の画像から数百個にも及ぶ「意味のある領域」と「説明文」のペアを候補として生成する．図 3 の②にその例を示す（可視化のため，スコア上位 50 件のみを表示している）．

第二段階では，第一段階で生成された無秩序な候補群の中から，提示に適した記述を抽出する．具体的には，画像を 3 × 3 のグリッドで覆い，9 つの各グリッドセルについて，第一段階で生成された全ての候補領域との空間的な重複度を IoU（Intersection over Union）を用いて算出する．IoU は，2 つのバウンディングボックスの重なり具合を 0 から 1 の値で評価する指標であり，2 つの領域の共通部分の面積を，和集合の面積で除算することで求められる．これを数式で表すと式 (1) の通りである．

$$IoU = \frac{Area(B_{grid} \cap B_{candidate})}{Area(B_{grid} \cup B_{candidate})} \tag{1}$$

式 (1) において，B_{grid} はグリッドセル，$B_{candidate}$ は候補領域のバウンディングボックスを示す．図 3 の③に計算の概念図を示す．各グリッドセル（図中の青枠）に対し，IoU が最大となる候補領域（図中の赤枠）に紐づく説明文を，そのセルを代表する最終的な記述として採用する．図 3 の④は，この選択プロセスを経た最終出力を示しており，各グリッドセルに対して選択された説明文と，その根拠となった最大 IoU 値が列挙されている．

この 2 段階のプロセスにより，本手法は最終的に 9 つの構造化された説明文を出力する．最終的な出力形式は部分領域記述法と同様であるが，各説明文が単純な矩形の切り抜きではなく，意味のある物体や部分を捉えた領域に基づいている点で質的に異なる．本手法は，Dense Captioning が持つ「意味的な詳細さ」と，グリッド構造による「提示の明確さ」の両立を実現するものである．

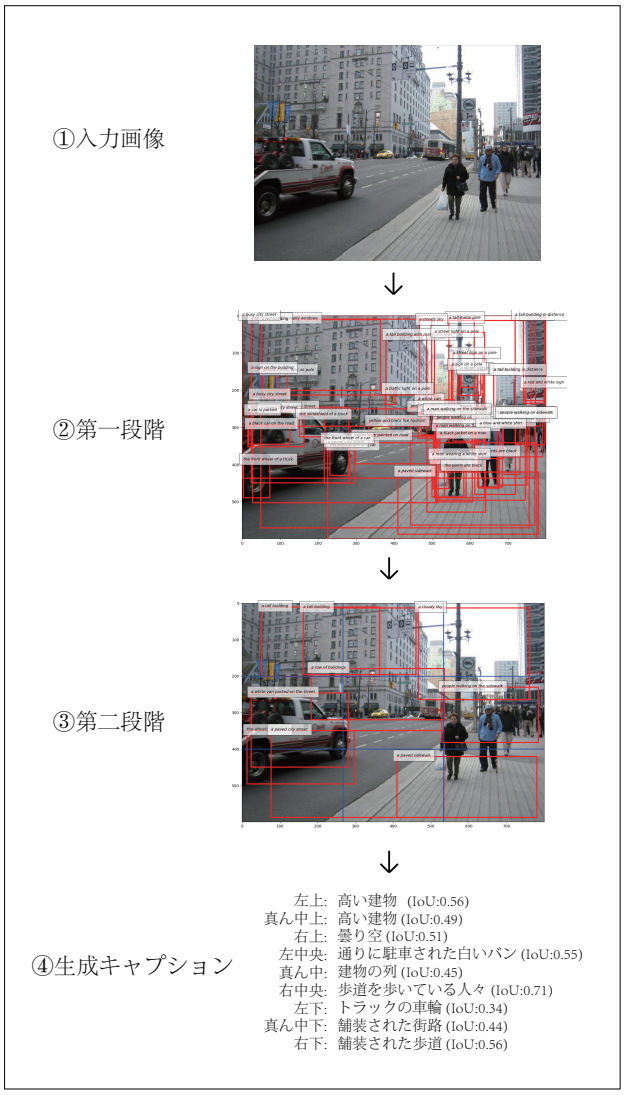


図 3: 提案手法 2 の実行例．画像は Visual Genome データセットから引用．

4 評価実験

4.1 実験の概要

提案した部分領域記述法と重ね領域最大法の有効性を評価するため，ユーザ参加型の比較実験を実施した．従来的なアプローチである単一の包括的な説明文を生成する手法をベースラインとして設定し，これら 3 つの手法が生成した説明文の理解しやすさを多角的に評価する．本実験では，20 名の参加者を募集し，提案手法により生成された画像説明文のみを頼りに，元の画像を想像してスケッチを描いてもらうタスクを課した．

実験評価として，客観的評価と主観的評価の 2 つの指標を記録した．客観的評価では，ユーザの「感知と理解」の度合いを測定するため，参加者が描いたスケッチと元の画像を大規模言語モデルに入力し，両者の類似度スコアを算出させた．これに加え，主観的評価として，参加者自身が各説明文からどれだけ鮮明に画像を想像できたかを「想像のしやすさ」として 5 段階で評価した．このように，本研究では客観的な大規模言語モデルによる類似度スコアと主観的なユーザ評価を組み合わせることで，各記述法がユーザの心的イメージ構築に与える影響を多角的に分析した．

4.2 実験環境・設定

4.2.1 実験用データ

実験には 10 カテゴリのデータを使用した．これらのカテゴリは，提案手法 2（重ね領域最大法）の基盤である Visual Genome データセット [13, 15] と，提案手法 1（部分領域記述法）の基盤である CLIP ベースのモデル [12, 16] が得意とするデータセットの 2 つのグループから抽出された．

具体的には，Visual Genome グループからは，「スキー」や「シマウマ」といった同データセットの主要カテゴリ（人物，スポーツ，動物）や，「信号機」のような一般的な物体認識能力を評価するカテゴリが選ばれた．CLIP ベースのグループからは，「車」や「料理」といった専門ドメインにおけるきめ細かい分類能力や，「風景」や「日常」のような広範なシーン認識・汎用性能を評価するカテゴリが選定された．

カテゴリは，単一の被写体を含む画像，複数のオブジェクトを含む複雑なシーン，および様々な専門ドメインを網羅し，内容の多様性を確保するように意図的に選定された．

4.2.2 実験用インタフェース

目の不自由なユーザが画像を直接見ることができないシナリオをシミュレートするため，我々は画像を表示せずに説明文を提供する UI を開発した（図 4 参照）．この UI は，画面上部の情報提示エリア，下部の描画エリア，そして右側の制御・評価エリアという 3 つの主要な領域で構成されている．参加者はこの UI を通じて各手法が生成した説明文を受け取り，タスクを遂行した．

情報提示の形式は，制御エリアの「方法」ボタンによって選択された手法に応じて変化する．「clip_1」（ベースライン）が選択された場合，画像全体を要約する単一の説明文が右側のテキストボックスに直接表示される．一方，「clip_2」（部分領域記述法）または「densecap」（重ね領域最大法）が選択された場合は，左上のエリアがインタラクティブな 3 × 3 グリッドとして機能する．参加者がこのグリッドのいずれかの区画をクリックす

図 4: 提案システム用の実験インタフェース

ると，その特定領域に対応する説明文が右側のテキストボックスに表示される．

参加者は提示された説明文に基づき，下部の「描画エリア」にスケッチを描く．描画されるスケッチの解像度は元の画像と完全に一致するよう制御され，後の客観評価（LLM による類似度スコア算出）の公平性を確保している．描画完了後，参加者は右下の「評価」ボタンを使い，説明文の「想像のしやすさ」を 5 段階で評価し，結果を送信する．

4.3 実験手順

本研究は，以下の手順で実験を実施した．

1. 合計 100 枚の画像に対し，ベースライン手法（単一記述法），提案手法 1（部分領域記述法），提案手法 2（重ね領域最大法）の 3 つの手法で説明文を生成し，合計 300 の説明文セットを作成した．
2. これら 300 個の説明文セットをランダムに 20 組へ均等に分割した．各組には 15 件の説明文（3 手法× 5 画像分）が含まれており，組番号は 1〜20 である．
3. 実験には 20 名の参加者を募集し，各参加者に 1 から 20 のいずれかの組番号を割り当てた．
4. 各参加者に対し，割り当てられた組に含まれる 15 件の説明文を 1 つずつ提示し，それぞれの説明文に基づいて簡易的なスケッチを描画してもらった．
5. 加えて，参加者に対し「説明文から想像しやすいかどうか（画面構成の理解容易性）」について 5 段階で主観的に評価してもらった．

4.4 評価方法

4.4.1 客観評価

我々は大規模言語モデル（LLM）を用いた独自の評価フレームワークを構築した．具体的には，まず参加者がテキスト記述のみに基づいて描いたスケッチを収集した．その後，LLM に対し，参加者のスケッチと元の参照画像との類似度スコアを算出させた．

この際，LLM にはプロンプトを与え，0 から 100 の間の類似度スコアを出力させた．このスコアを，本研究における「感知と理解」の客観的指標として採用した．

生成 AI の結果は，実行ごとに割り当てられるシード値によって僅かに変動する可能性があることを考慮し，本研究では詳細な採点基準を設計しただけでなく，各画像ペアに対する評価を 20 回繰り返した．そして，これらの評点の平均値を，最終的な「理解度」の客観スコアとして採用した．

4.4.2 主観評価

評価指標として「想像のしやすさ」を設定し，これは「説明文からどれだけ容易に画面全体の様子を想像できるか，また，構図を理解しやすいか」と定義される．参加者には，各説明文を評価した後，この「想像のしやすさ」について 1（非常に想像しにくい）から 5（非常に想像しやすい）までの 5 段階で評定してもらった．収集された評定値は，手法ごとに平均値や分散などの統計量を算出し，その分析を通じて各アプローチの優劣を比較した．

4.5 実験結果

4.5.1 描画結果の比較

各手法が生成した説明文から参加者が実際にどのような画像を思い描いたかを探るため，描画されたスケッチの典型的な例を提示する．図 5 は，同一の元画像に対し，3 つの異なる手法で生成された説明文に基づいて描かれたスケッチの比較例である．

4.5.2 客観評価（LLM による類似度スコア）の集計結果

全画像を通した手法ごとの全体的な性能と，画像カテゴリごとの性能を詳細に分析するため，それぞれ統計量を算出した．表 1 は，全カテゴリを対象とした各手法の平均スコアと分散を示す．重ね領域最大法が最も高い平均スコア（30.83）を示し，ベースラインは最も低い（24.48）．表 2 は，各カテゴリごとに手法別の

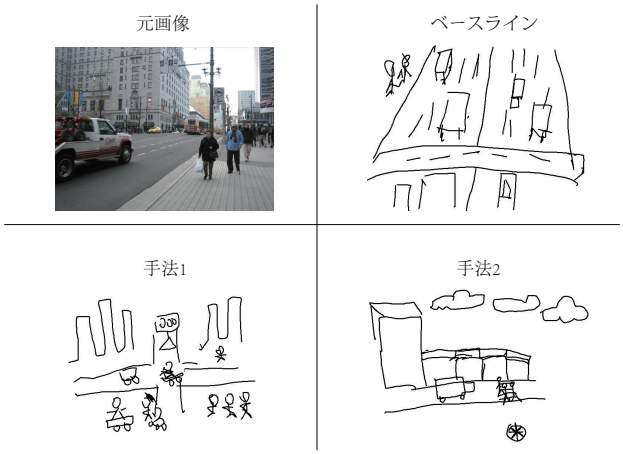


図 5: 描画されたスケッチの比較例．画像は Visual Genome データセットから引用．

表 1: 全カテゴリにおける各手法の平均スコアと分散

手法	平均スコア	分散
ベースライン	24.48	93.22
部分領域記述法	25.12	97.75
重ね領域最大法	**30.83**	117.81

平均スコアを示している．「06 車」カテゴリでは手法 2 が最も高いスコア（37.875）を記録し，「08 料理」ではベースラインが最も高い（21.390）など，カテゴリによって異なる傾向が見られる．

各手法間に統計的に有意なスコア差が存在するかを検証するため，全体およびカテゴリ別に対して一元配置分散分析（ANOVA）を実施した．全体では有意差は観察されなかった．カテゴリ別では，「04 シマウマ」と「05 標識」において，p 値がそれぞれ 0.022 と 0.004 と有意水準（$p < 0.05$）を下回り，統計的に有意な差が認められた（表 3 参照）．

表 2: カテゴリごとの各手法の平均スコア

カテゴリ	ベースライン	手法 1	手法 2
01 スキー	32.965	32.230	**36.610**
02 野球	25.135	33.285	**37.260**
03 信号機	27.850	34.965	**35.185**
04 シマウマ	34.180	15.130	**34.325**
05 標識	14.030	32.550	**37.510**
06 車	21.360	15.565	**37.875**
07 風景	17.435	18.960	**26.820**
08 料理	**21.390**	14.485	7.335
09 部屋	20.695	**30.865**	23.750
10 日常	29.795	23.165	**31.670**

表 3: 各カテゴリにおける 3 手法間の客観評価スコアの一元配置分散分析（ANOVA）結果

カテゴリ	F 値 (F-statistic)	p 値 (p-value)
全データ	2.99	0.052
01 スキー	0.15	0.863
02 野球	0.68	0.517
03 信号機	0.42	0.663
04 シマウマ	**4.44**	**0.022**
05 標識	**6.84**	**0.004**
06 車	2.55	0.097
07 風景	0.95	0.401
08 料理	1.63	0.214
09 部屋	0.56	0.579
10 日常	0.45	0.641

表 4: カテゴリ別の主観評価スコア，および全体の平均と分散

カテゴリ	ベースライン	手法 1	手法 2
01 スキー	3.3	3.6	4.0
02 野球	3.0	2.9	4.1
03 信号機	2.8	3.3	3.8
04 シマウマ	4.0	2.8	4.5
05 標識	3.0	3.1	3.7
06 車	3.6	3.6	3.9
07 風景	2.8	3.6	3.6
08 料理	2.7	2.2	4.1
09 部屋	2.6	3.7	3.3
10 日常	4.1	3.9	3.8
平均スコア	**3.19**	**3.27**	**3.88**
分散	**1.27**	**1.18**	**0.89**

4.5.3 主観評価（想像しやすさ）の集計結果

次に，「想像のしやすさ」に関する主観評価の集計結果について述べる．表 4 は，この評価の詳細な結果を示す．各カテゴリの平均スコアと共に，各手法の全体的な平均スコアと分散が示されている．全体として，手法 2（重ね領域最大法）が最も高い平均スコア（3.88）と最も低い分散（0.89）を達成した．カテゴリ別に見ると，「04 シマウマ」や「08 料理」で手法 2 の評価が特に高く，一方で「09 部屋」では手法 1 が最も高い評価を受けた．

5 実験結果に対する考察

本章では，実験結果に基づき，各手法の有効性と画像カテゴリの特性が与える影響について多角的な考察を行う．

5.1 手法ごとの全体的な有効性に関する考察

まず，全カテゴリを総合した際の各手法の有効性について考察する．客観評価において，提案手法，特に重ね領域最大法（手法 2）が優れている傾向が観察された．表 1 が示す通り，重ね領域最大法の平均スコアは 30.83 であり，部分領域記述法（手法 1）の 25.12，ベースラインの 24.48 を明確に上回った．この差について一元配置分散分析（ANOVA）で検定したところ，有意差はなかった．本実験の参加者数 (N=20) が，全体差を検出する上での統計的検出力に影響した可能性が考えられる．

一方で，この高い平均スコアが示す潜在的な優位性の傾向は，主観評価の結果によっても裏付けられた．表 4 を見ると，重ね領域最大法の「想像のしやすさ」に関する平均スコアは 3.88 であり，ベースライン（3.19）および部分領域記述法（3.27）を大幅に上回った．この結果は，ユーザが重ね領域最大法から得た情報を，心的イメージの構築に最も容易だと感じたことを示している．

客観・主観両方の評価結果を総合すると，単一の記述よりも複数の構造化された記述を提供する方がユーザの画像理解を効果的に支援するという傾向が観察された．その中でも，意味のある領域を的確に捉えて記述する重ね領域最大法が最も有望なアプローチとして際立っており，本提案システムの有効性を示している．

5.2 画像カテゴリの特性による影響の考察

次に，結果をカテゴリ別に詳細に分析し，最適な記述法が画像の特性に強く依存していることを明らかにする．第一に，単一の明確な被写体が画像の大部分を占める場合，ベースライン手法が有効な場合がある．例えば，表 2 の「08 料理」カテゴリでは，ベースラインの平均スコアが 21.390 と最も高かった．ただし，この差は統計的に有意ではなかった．この傾向は，画像が単一の被写体と単純な情報を含む場合，画像を分割するアプローチが不必要に情報を断片化させ，単一の包括的な記述文の方が効率的である可能性を示唆している．

第二に，単一被写体であっても，機械的なグリッド分割は有効に機能しないことが示された．「04 シマウマ」カテゴリでは，手法間に統計的に有意な差が検出された（$p = 0.022 < 0.05$）．表 2 を見ると，部分領域記述法のスコア（15.130）は，ベースライン（34.180）および重ね領域最大法（34.325）より著しく低い．これは，グリッド分割がゼブラという被写体を不自然に分断したためと考えられる．対照的に，ゼブラ全体を単一の意味のある領域として捉えた重ね領域最大法と，画像全体を記述したベースラインは，共に高い評価を受けた．

第三に，画像内に比較的小さくとも重要な要素が含まれる場合，重ね領域最大法が圧倒的に優位であった．「05 標識」カテゴリでは，手法間に有意な差が観察され（$p = 0.004 < 0.05$），重ね領域最大法のスコア（37.510）はベースラインの（14.030）を遥かに上回った（表 2）．これは，DenseCap ベースの手法が，道路標識という特定の重要領域を正確に検出し，その内容を詳細に記述したことが，ユーザの理解に直接貢献したためと考えられる．

最後に，構造的なシーンでは部分領域記述法が有効な場合もあった．「09 部屋」カテゴリでは，部分領域記述法の平均スコアが 30.865 と最も高かった（表 2）．この差は統計的に有意ではなかったが，この傾向は，部屋のような人工的な環境では，「左上・中央・右下」といった単純なグリッド構造が，空間全体のレイアウトを直感的に伝える上で有効であった可能性を示唆している．

これらのカテゴリ別の詳細な考察は，単一の画像記述アプローチですべてのカテゴリに最適に対応することが困難であることを示している．したがって，画像の構成や内容に応じて最適な記述戦略を動的に選択できる，柔軟な画像記述システムが有効であろうことが確認された．

6 結論

本研究は，単一の包括的な記述文では目の不自由なユーザが複雑な画像を理解するには不十分であるという課題に取り組んだ．画像を複数領域に分割して説明する「サブ画像記述」アプローチを提案し，その有効性を検証した．具体的には，ベースラインである単一記述法，部分領域記述法，および重ね領域最大法の 3 つの手法を，参加者が説明に基づいてスケッチを描くという独自の実験手法を用いて比較した．客観・主観評価による実験結果は，複数の記述を提供する提案手法，特に重ね領域最大法がベースラインを上回る傾向を示した．同時に，単一の被写体が中心の画像ではベースラインがより効果的な場合があるなど，最適な記述法が画像の特性に強く依存することも明らかになった．本研究の貢献は，「サブ画像記述」アプローチの有効性を実証し，ユーザの理解度を定量化する新たな評価方法を提示した点にある．

今後の改良点として，領域分割の最適化と説明の個別化が挙げられる．具体的には，画像の複雑さに応じて領域を動的に決定する適応的セグメンテーションや，ユーザが関心領域を指定する機能の導入を計画している．また，ユーザの知識レベルや好みに応じて説明を調整し，リアルタイムのフィードバックに基づいて内容を動的に変更する機能の開発も目指す．これらの機

能強化は，言語モデルのパラメータ調整や，より表現豊かなデータセットでのモデルの再訓練によって実現可能である．さらに，本実験 (N=20) では $p = 0.052$ という有意傾向に留まった全体差を明確に検証するため，提案システムの有効性をより強固に検証するため，今後はより多くの参加者による評価を実施する予定である．

参考文献

[1] GBD 2019 Blindness and Vision Impairment Collaborators: Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study, *The Lancet Global Health*, Vol. 9, No. 2, pp. e130–e143 (2021)

[2] Vision Loss Expert Group of the Global Burden of Disease Study.: Publisher Correction: Global estimates on the number of people blind or visually impaired by cataract: a meta-analysis from 2000 to 2020, *Eye*, Vol. 38, No. 11, pp. 2229 (2024)

[3] Hou, W., Riccò, D.: Accessible Design for Museums: A Systematic Review on Multisensory Experience Based on Digital Technology, *Advances in Design and Digital Communication V*, Vol. 51, pp. 282–298 (2025)

[4] Cavazos Quero, L., Iranzo Bartolomé, J., Cho, J.: Accessible visual artworks for blind and visually impaired people: comparing a multimodal approach with tactile graphics, *Electronics*, Vol. 10, No. 3, pp. 297 (2021)

[5] Petrie, H.: Crowdsourcing descriptions of visual works of art for blind and partially sighted people, *Diss. York* (2023)

[6] Li, J., Li, D., Savarese, S., Hoi, S.: BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models, *arXiv preprint arXiv:2301.12597* (2023)

[7] Liu, H., Li, C., Wu, Q., Lee, Y. J.: Visual Instruction Tuning, *arXiv preprint arXiv:2304.08485* (2023)

[8] Fernando, S., Ndukwe, C., Virdee, B., Djemai, R.: Image recognition tools for blind and visually impaired users: An emphasis on the design considerations, *ACM Transactions on Accessible Computing*, Vol. 18, No. 1, pp. 1–21 (2025)

[9] Li, F. M., Zhang, L., Bandukda, M., Stangl, A., Shinohara, K., Findlater, L., Carrington, P.: Understanding visual arts experiences of blind people, *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023)

[10] Doore, S. A., Istrati, D., Xu, C., Qiu, Y., Sarrazin, A., Giudice, N. A.: Images, Words, and Imagination: Accessible Descriptions to Support Blind and Low Vision Art Exploration and Engagement, *Journal of Imaging*, Vol. 10, No. 1, pp. 26 (2024)

[11] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision, *Proceedings of the 38th International Conference on Machine Learning*, Vol. 139, pp. 8748–8763 (2021)

[12] Mokady, R., Hertz, A., Bermano, A. H.: Clipcap: Clip prefix for image captioning, *arXiv preprint arXiv:2111.09734* (2021)

[13] Johnson, J., Karpathy, A., Fei-Fei, L.: Densecap: Fully convolutional localization networks for dense captioning, *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016)

[14] Delloul, K., Larabi, S.: Towards Real Time Egocentric Segment Captioning for The Blind and Visually Impaired in RGB-D Theatre Images, *arXiv preprint arXiv:2308.13892* (2023)

[15] Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L.-J., Shamma, D. A., Bernstein, M. S., Li, F.-F.: Visual genome: Connecting language and vision using crowdsourced dense image annotations, *International Journal of Computer Vision*, Vol. 123, No. 1, pp. 32–73 (2017)

[16] Sharma, P., Ding, N., Goodman, S., Soricut, R.: Conceptual captions: A cleaned, hypernymed, image alt-text dataset for automatic image captioning, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2556–2565 (2018)

画像生成 AI を用いた読者を誘引する書籍表紙画像の生成

池田 諒真 [1] Shan Junjie[1] 安尾 萌 [2] 西原 陽子 [1]
Ryoma Ikeda[1] Junjie Shan[1] Megumi Yasuo[2] Yoko Nishihara[1]

[1] 立命館大学情報理工学部
[1] College of Information Science and Engineering, Ritsumeikan University
[2] 立命館グローバル・イノベーション研究機構
[2] Ritsumeikan Global Innovation Research Organization

Abstract: 書籍の表紙は，読者に与える第一印象を形成する．既存研究では，書籍表紙のデザインは読者の興味や選択に影響を与えることが明らかになった．しかし，表紙画像に書籍のタイトルやカテゴリがどのように反映されれば，読者を誘引する表紙画像となるかはまだ不明である．本研究では，テキストから画像を生成する AI モデルを活用し，書籍の説明文から表紙画像を生成するシステムを構築した．このシステムを用い，カテゴリ別の書籍に対して，異なるプロンプト構成によって生成された表紙画像が読者を誘引するかについて主観実験を行い評価した．実験の結果，書籍の説明文にある名詞を全て画像生成 AI のプロンプトとした表紙画像が最も読者を誘引した．

1 はじめに

商品パッケージのデザインは，消費者に与える第一印象を形成し，その購買意欲に大きな影響を及ぼす [Bloch 95, Silayoi 07]．書籍の表紙もまた「書籍商品のパッケージ」として，読者に第一印象を与え，手に取るかどうかの判断を左右する重要な要素である [Hagtvedt 08, Jian 19]．既存の研究では，表紙に描かれたポジティブな感情を喚起する写真やイラストが， 結果として購買意欲を高める可能性が示唆されている [Liu 17]．しかし，表紙画像の具体的なデザイン属性と，読者の興味との関連性については，十分な分析が行われていないのが現状である [Orth 08]． さらに，これまでの書籍表紙デザインは専門デザイナーの感覚や経験に大きく依存してきたため，統一的な基準による定量的・定性的な分析は困難であった [Bruce 97].

本研究では，画像生成 AI を用いて書籍の説明文から表紙画像を自動生成し，どのような属性をプロンプトの構成とすると**書籍表紙の誘引力**を高める表紙画像を生成できるかを明らかにすることを目的とする．近年注目されている入力テキストから画像を生成する技術 [Ramesh 21, Rombach 22] を用いて，書籍の説明文から表紙画像を生成するシステムを構築した． 本研究で達成した主な点は以下の通りである．

- 画像生成 AI を利用し，書籍の説明文から表紙画像を生成するシステムを構築した．
- 書籍の説明文にある単語の品詞と重要度に基づき，画像生成 AI に入力するプロンプトの内容を構成する 4 つの手法を提案し，それぞれ表紙画像を生成した．
- 主観実験を通じて，4 つの手法を用いて生成された，18 種類の書籍カテゴリ別の表紙画像の誘引力を評価し，カテゴリ別の各提案手法で生成した表紙画像が，どれだけ読者を誘引するかを分析した．

2 関連研究

2.1 認知に関する既存研究

Coherence Principle[Mayer 05] は，視覚情報とテキスト情報が意味的に一致しているほど，学習者の認知的な負荷が下がり，内容の理解が促進され，結果として好意的な評価につながるとしている．この原則は，情報の整合性が人間の認知プロセスにおいて重要な役割を果たすことを示唆している．

2.2 心理学に関する既存研究

Processing Fluency Theory[Reber 04] は，人間は知覚的に処理しやすい，つまりスムーズに理解できる情報を好む傾向があることを示している．これは,「分かりやすいものは心地よい」という直感的な感覚を理論的に説明するものである．

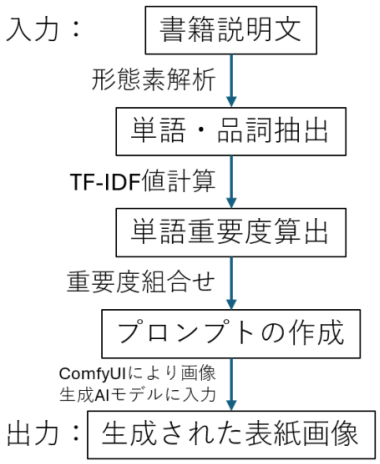


図 1: 提案システムのフロー図

2.3 書籍表紙・商品パッケージに関する既存研究

商品パッケージのデザインが消費者の購買意欲に与える影響については，Bloch ら [Bloch 95] や Silayoi ら [Silayoi 07] をはじめ，多くの研究が行われてきた．書籍の表紙も同様に「パッケージ」として，読者の選択に影響を与える重要な要素である [Hagtvedt 08]．Gudinavičius ら [Gudinavičius 17] は，読者の選択における表紙の役割を分析し，その重要性を指摘している．特にオンライン書店の普及により，読者は中身を読む前に表紙画像だけで購入を判断する機会が増えており [Jiang 13]，Park ら [Park 23] もオンライン環境での表紙の役割について分析している．しかし，表紙画像の具体的なデザイン属性と，読者の興味との関連性については，十分な分析が行われていないのが現状である．

3 書籍表紙画像の生成手法

本研究では，書籍の説明文から読者を誘引する表紙画像の生成を目的とするため，画像生成 AI に入力するプロンプトの作成により表紙画像の生成を調整している．**図 1** に示す提案システムのフロー図に基づき，入力された「書籍説明文」に対し，形態素解析と TF-IDF 値の計算を行うことで，プロンプトを作成するための「単語重要度」を算出する．次に，抽出された単語や重要度の組合せにより，プロンプトの作成手法を提案する．作成したプロンプトを画像生成 AI モデルに入力し，表紙画像を生成する．本章では，これらの手順について詳細に述べる．

3.1 説明文の単語抽出

本研究では，画像生成 AI への指示となるプロンプトを作成するため，形態素解析により書籍の説明文から単語と品詞の抽出を行う．形態素解析エンジンには MeCab[1] を使用し，新しい単語に対応するため追加辞書 NEologd[2] を用いる．画像生成 AI のプロンプトには名詞で構成されており，また書籍の説明文は内容を客観的に伝えるために主観的な記述（形容詞，副詞など）を避けて名詞中心の構成となる傾向があることを考慮する． 本研究では形態素解析により抽出された単語の中から名詞をプロンプトの要素として採用した．抽出された名詞のうち，プロンプトとして意味内容が希薄となる代名詞， 非自立， 数， 接尾， 副詞可能， 助動詞語幹， 接続詞的といった単語を除外し，データのクリーニングを行う．

3.2 単語重要度の算出

プロンプトに使用する単語の重要度を決めるため，形態素解析後の各単語に対して TF-IDF 値を計算する．これにより，各書籍の内容を特徴づける重要な単語が数値化される．IDF 値の作成には，最新版の日本語 Wikipedia の全記事データ[3] を基に構築されたテキストコーパスを使用した．IDF 値について，ユニーク単語数が 3,725,459(総単語数が 720,205,244)， 総文書数 (記事数) が 2,374,911 となっている．**表 1** に，書籍説明文の入力と抽出された名詞のサンプル例を示す．

3.3 プロンプトの生成

画像生成 AI において，プロンプトは大きく分けて「Positive」プロンプトと「Negative」プロンプトの 2 種類があり，それぞれが生成される画像の内容を制御する上で重要な役割を果たす．Positive プロンプトは「生成したい要素」や「画像に含めたい属性」を AI に指示するものであり，一方 Negative プロンプトは「生成したくない要素」や「除外したい属性」（例えば，奇形になりやすい文字，不完全・歪みな画像など ）を指示している． 本研究では，説明文から抽出された各単語と重要度の組み合わせにより，以下の 4 種類のプロンプト作成方法を提案する．本研究では，3.1 で抽出した名詞を英訳した上で，書籍の内容を反映させるための 4 つの手法を比較した．

- **手法 1（全名詞 Positive）：** 書籍説明文から抽出された名詞を全て平等（重み 1.0）に Positive の

[1] https://taku910.github.io/mecab/
[2] https://github.com/neologd/mecab-ipadic-neologd
[3] https://dumps.wikimedia.org/jawiki/latest/jawiki-latest-pages-articles.xml.bz2

表 1: 形態素解析による単語抽出の具体例

説明文入力
DNA や生体情報の収集，顔や声を常時スキャンする街頭センサー，移動・購買履歴を自動で吸い上げる端末．すべてのデータは中央 AI に集約され，アルゴリズムが個人ごとに「危険度スコア」を算出する仕組みになっていた．

抽出単語

単語 (上位 20 件)	TF-IDF 値
吸い上げる	9.7826
危険度	8.4344
スキャン	7.5788
購買	7.3223
街頭	7.3178
履歴	6.6423
アルゴリズム	6.6275
生体	6.6052
センサー	6.6024
AI	6.5507
算出	6.4946
DNA	6.3542
常時	6.3176
端末	6.1619
スコア	6.0428
集約	6.0317
仕組み	5.5566
自動	5.2906
収集	5.2541
顔	4.6522

内容としてプロンプトを作成する．これは，書籍説明文にある単語は，書籍の内容を表現するには全て重要であると考えられるためである．

- **手法 2（Positive/Negative 分割）：** 書籍説明文から抽出された名詞を重要度（TF-IDF 値）により並び替え， 重要度の上位半分を Positive プロンプト， 重要度の下位半分を Negative プロンプトとしてプロンプトを作成する． 各部分に入力する単語の重みは全て 1.0 にする．これは，書籍説明文にある単語は「独自性のある単語」と「一般的な単語」が含まれ， 独自性のある単語を強調させて，一般的な単語を避けるために考えられた．

- **手法 3（上位 10 件）：** 書籍説明文から抽出された名詞を重要度（TF-IDF 値）により並び替え，重要度上位 10 件の名詞のみを Positive プロンプトとしてプロンプトを作成する．なお，入力する各単語の重みは全て同じ（重み 1.0）である．これは，手法 2 の考え方と同じく，書籍の内容に繋がる単語を強調する．

- **手法 4（重み付け）：** 書籍説明文から抽出された名詞を重要度（TF-IDF 値）により並び替え，重要度上位 2 件の名詞に 1.15 倍，3 位から 10 位の名詞に 1.1 倍，残る単語に 1.0 倍の重みを付けて，Positive プロンプトとして入力してプロンプトを作成する．この処理は，書籍説明文にある単語が書籍の内容を表現する上で重要なものであると考えられるために行うものである．重み付けの基準については，画像生成 AI の重み付けに関する検証記事 4 を参考に重みを決定し，Positive プロンプトとして入力する．

主流なオープンソースの画像生成 AI モデルは， 英語の入力しか受け入れられないので，各手法により抽出された単語を英訳した上で，画像生成モデルに入力する．その上，全提案手法において，表紙画像の生成を指示する「(cover image:1.2), key visual」の Positive のプロンプトと，奇形・劣化の画像を抑制する「text, title, watermark, cropped, out of frame」の Negative のプロンプトを入力している．本モデルにおいて最も高品質かつ安定した画像生成が期待できる標準解像度であることから，生成される画像の解像度は 1024 × 1024 とした．

表 2 は各手法により作成されたプロンプトの構成例を示している．本研究では，ComfyUI[4] と Stable Diffusion 3.5[5] モデルを利用して表紙画像の生成を行う．**図 2** に各手法の入力プロンプトから生成された表紙画像の例を示す．

4 評価実験

4.1 実験方法と評価指標

生成した表紙画像が読者の興味をどの程度誘引するかを評価するため，主観実験により評価した．書籍の購入決定には，価格，著者，レビュー，販売環境など様々な外的要因が複雑に絡み合う．そのため本研究では，これらの外的要因を捨象し，表紙デザインそのものが持つ，読者の**「興味を惹き，手に取らせる力（誘引力）」**に焦点を当てることとした．この指標を，本稿では便宜上**「魅力度」**と呼ぶこととする．ただし，評価実験の実施にあたっては，この「**魅力度**」という概念を被験者により具体的に理解してもらうため，指示文や評価基準の説明において**「書籍の表紙としてどの程度購買意欲が湧くか」**という言葉を用いた． 評価実験で使用した UI を図 3 と図 4 に示す． 図 3 は 4 つの表紙画像の内 1 つを選択した際の， 「書籍タイトルと生成された表紙画像」に対する 5 段階評価の画面である． 図 4 は書籍タイトルと各手法により生成された表紙画像を表示しており（左から順に手法 1，手法 2，手

[4]https://github.com/comfyanonymous/ComfyUI
[5]https://stability.ai/stable-diffusion-3

図 2: **表 2** に示した各手法で作成したプロンプトにより生成された表紙画像例（左から手法 1，2，3，4）

表 2: 各手法におけるプロンプト構成例

手法	Positive プロンプト	Negative プロンプト
手法 1	(cover image:1.2), key visual, Monitoring, data, City, Risk level, Vivid, cage, scan, Near Future, Tech, Buy, street, back side, history, algorithm, living organisms, sensor, AI, calculation, technology, transparent, DNA, Terminal, Score, Aggregation, structure, daily, future, automatic, collection, Major, escape, net, digital, girl, shaft, huge, the same, Device, face, city, residents, story, Domination, society, voice, central, move, system, personal, people, figure, enterprise, information, development, record	text, title, watermark, cropped, out of frame
手法 2	(cover image:1.2), key visual, Monitoring, data, City, Risk level, Vivid, cage, scan, Near Future, Tech, Buy, street, back side, history, algorithm, living organisms, sensor, AI, calculation, technology, transparent, DNA, Terminal, Score, Aggregation, structure, daily, future, automatic	text, title, watermark, cropped, out of frame, collection, Major, escape, net, digital, girl, shaft, huge, the same, Device, face, city, residents, story, Domination, society, voice, central, move, system, personal, people, figure, enterprise, information, development, record
手法 3	(cover image:1.2), key visual, Monitoring, data, City, Risk level, Vivid, cage, scan, Near Future, Tech, Buy	text, title, watermark, cropped, out of frame
手法 4	(cover image:1.2), key visual, (Monitoring:1.15), (data:1.15), (City:1.1), (Risk level:1.1), (Vivid:1.1), (cage:1.1), (scan:1.1), (Near Future:1.1), (Tech:1.1), (Buy:1.1), street, back side, history, algorithm, living organisms, sensor, AI, calculation, technology, transparent, DNA, Terminal, Score, Aggregation, structure, daily, future, automatic, collection, Major, escape, net, digital, girl, shaft, huge, the same, Device, face, city, residents, story, Domination, society, voice, central, move, system, personal, people, figure, enterprise, information, development, record	text, title, watermark, cropped, out of frame

図 3: 「書籍タイトルと生成された表紙画像のペア」に対する主観的購買意欲の評価画面の 1 例

法 3，手法 4），5 段階評価後の最も「購買意欲が湧く」画像を選択する画面である．

実験は**図 3** および**図 4** の UI を用いて行われ，被験者には以下の手順で評価を依頼した．

1. 基本情報（性別，年齢など）を入力させる．
2. ランダムに 1 つの書籍タイトルが選択され，4 つの手法により生成された表紙画像を表示する．
3. 被験者に，**図 3** に示すように，表示された「書籍タイトルと生成された表紙画像のペア」に対して，他の画像と比較せず，「書籍の表紙としてどの程度購買意欲が湧くか」という基準で 5 段階評価する（絶対評価）．
4. 被験者に，**図 4** に示すように，4 つの画像の中から，「最も購買意欲が湧く」と感じた画像を 1 つ選択させる（相対評価）．

被験者には，評価に正解はなく，自身の感覚に基づいて正直に回答するよう教示した．

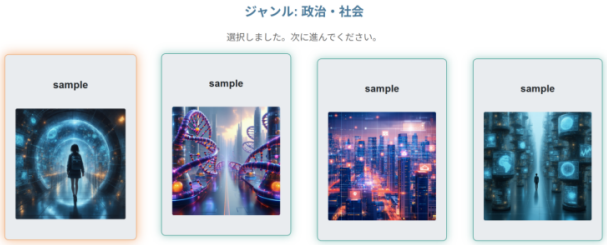


図 4: 書籍タイトルと各手法により生成された画像の表示画面（左から手法 1，手法 2，手法 3，手法 4 の順で書籍タイトルと生成された表紙画像のペア）

表 3: 実験に用いた書籍カテゴリとタイトル数

カテゴリ	タイトル数
アート・建築・デザイン	10
コミックス	10
サイエンス・テクノロジー	10
スポーツ・アウトドア	10
ノンフィクション	10
ビジネス・経済	10
音楽	10
絵本・児童書	10
教育・自己啓発	10
芸能・エンターテインメント	10
事典・年鑑・本・ことば	10
趣味・実用	10
人文・思想・宗教	10
政治・社会	10
文芸・評論	10
暮らし・健康・料理	10
旅行・紀行	10
歴史・地理	10
合計	**180**

4.2 実験データ

本実験では，**表 3** に示す 18 カテゴリから各 10 タイトル，計 180 タイトルの書籍データを使用した．実験参加者は 34 名（男性 17 名， 女性 17 名， 年齢は 15〜24 歳）であった．

4.3 実験結果

4.3.1 全体の評価とカテゴリ別分析

実験全体の結果をまとめる．**表 4** は，4 つの手法の「**魅力度**」に関する 5 段階評価の平均点と標準偏差 (SD) をカテゴリ別に示し，最終行に全体の平均点と SD を加えたものである．**表 4** の全体平均では全ての名詞を Positive プロンプトに用いる**手法 1** が**3.00**点と最も高く，次いで重要度の順に単語に重み付けを行う手法 4 が 2.91 点となった．標準偏差は全体平均で 1.12 から 1.18 の範囲にあり，各手法ともに評価には個人差があることが示された．特に手法 1 は平均点が最も高い一方で，標準偏差は 1.12 と最も小さく，比較的評価が安定していた．**表 4** のカテゴリ別に見ると，「音楽」（**3.27** 点）や「ビジネス・経済」（**3.11**点）などで手法 1 の評価点が特に高い．一方で，「絵本・児童書」（**3.34**点），「事典・年鑑・本・ことば」（**3.13**点）など 4 つのジャンルでは手法 4 が最高評価となった．

表 5 は，各手法が最終的に「最も購買意欲が湧く」として選択された回数を示す．**表 5** の全体合計では**手法 1** が**1,040 件**（34.0%）と最も多く選択された．**表 5** のカテゴリ別に見ても，多くのジャンルで手法 1 の選択数が最も多い（例：「ノンフィクション」で 70 件，「ビジネス・経済」で 69 件）．一方で，「事典・年鑑・本・ことば」では手法 4 が 57 件と最も多く選択され，「絵本・児童書」でも手法 4 が 59 件と手法 1(66 件) に次いで多く選択された．

表 4: 手順 3「絶対評価」により得られたカテゴリ別および全体の 5 段階評価平均点と標準偏差 （最大値:**太字・下線**, SD: 標準偏差

ジャンル	手法 1(SD)	手法 2(SD)	手法 3(SD)	手法 4(SD)
アート・建築・デザイン	2.84 (1.21)	2.49 (1.12)	2.62 (1.09)	**2.88** (1.21)
コミックス	**3.03** (1.07)	2.46 (1.21)	2.52 (1.13)	2.98 (1.15)
サイエンス・テクノロジー	**3.12** (0.99)	2.41 (1.09)	2.61 (1.11)	2.91 (1.11)
スポーツ・アウトドア	**2.98** (1.12)	2.69 (1.24)	2.40 (1.16)	2.94 (1.24)
ノンフィクション	**3.01** (1.25)	2.48 (1.10)	2.55 (1.15)	2.75 (1.21)
ビジネス・経済	**3.11** (1.16)	2.49 (1.14)	2.49 (1.14)	3.02 (1.13)
音楽	**3.27** (1.09)	2.51 (1.03)	2.78 (1.12)	3.16 (1.18)
絵本・児童書	3.26 (1.09)	2.31 (1.01)	2.71 (1.17)	**3.34** (1.11)
教育・自己啓発	**2.98** (1.18)	2.65 (1.13)	2.42 (1.06)	2.50 (1.10)
芸能・エンターテインメント	**2.74** (1.15)	2.61 (1.04)	2.41 (1.16)	2.72 (1.19)
事典・年鑑・本・ことば	3.01 (1.10)	2.69 (1.10)	2.86 (1.13)	**3.13** (1.33)
趣味・実用	**2.85** (1.06)	2.39 (1.05)	2.67 (1.18)	2.76 (1.22)
人文・思想・宗教	**3.08** (1.05)	2.83 (1.10)	2.74 (1.15)	3.03 (1.20)
政治・社会	**2.96** (0.99)	2.75 (1.13)	2.70 (1.15)	2.71 (1.18)
文芸・評論	**2.86** (1.11)	2.62 (1.15)	2.35 (1.09)	2.84 (1.08)
暮らし・健康・料理	**2.94** (1.06)	2.75 (1.18)	2.89 (1.19)	2.78 (1.10)
旅行・紀行	2.84 (1.09)	2.67 (1.12)	2.64 (1.14)	**2.99** (1.11)
歴史・地理	**3.12** (1.11)	2.57 (1.13)	2.71 (1.13)	2.99 (1.17)
全体平均	**3.00** (1.12)	2.58 (1.13)	2.61 (1.15)	2.91 (1.18)
最大値の合計	**14**	0	0	4

表 5: 手順 4「相対評価」により得られたカテゴリ別の選択手法の分布（最大数:**太字・下線**, %: 割合）

ジャンル	手法 1(%)	手法 2(%)	手法 3(%)	手法 4(%)
アート・建築・デザイン	**57** (33.5%)	39 (22.9%)	28 (16.5%)	46 (27.1%)
コミックス	**53** (31.2%)	45 (26.5%)	22 (12.9%)	50 (29.4%)
サイエンス・テクノロジー	**69** (40.6%)	29 (17.1%)	35 (20.6%)	37 (21.8%)
スポーツ・アウトドア	44 (25.9%)	**54** (31.8%)	25 (14.7%)	47 (27.6%)
ノンフィクション	**70** (41.2%)	27 (15.9%)	36 (21.2%)	37 (21.8%)
ビジネス・経済	**69** (40.6%)	33 (19.4%)	26 (15.3%)	42 (24.7%)
音楽	**67** (39.4%)	19 (11.2%)	31 (18.2%)	53 (31.2%)
絵本・児童書	**66** (38.8%)	21 (12.4%)	24 (14.1%)	59 (34.7%)
教育・自己啓発	**63** (37.1%)	53 (31.2%)	24 (14.1%)	30 (17.6%)
芸能・エンターテインメント	**52** (30.6%)	48 (28.2%)	37 (21.8%)	33 (19.4%)
事典・年鑑・本・ことば	48 (28.2%)	26 (15.3%)	39 (22.9%)	**57** (33.5%)
趣味・実用	**51** (30.0%)	33 (19.4%)	42 (24.7%)	44 (25.9%)
人文・思想・宗教	**47** (27.6%)	**47** (27.6%)	35 (20.6%)	41 (24.1%)
政治・社会	**52** (30.6%)	43 (25.3%)	39 (22.9%)	36 (21.2%)
文芸・評論	**57** (33.5%)	41 (24.1%)	30 (17.6%)	42 (24.7%)
暮らし・健康・料理	**56** (32.9%)	42 (24.7%)	35 (20.6%)	37 (21.8%)
旅行・紀行	**56** (32.9%)	32 (18.8%)	33 (19.4%)	49 (28.8%)
歴史・地理	**63** (37.1%)	37 (21.8%)	35 (20.6%)	35 (20.6%)
全体合計	**1,040** (34.0%)	669 (21.9%)	576 (18.8%)	775 (25.3%)
最大数の合計	**16**	2	0	1

表 6: 手順 3「絶対評価」により得られた性別・カテゴリ別の 5 段階評価平均点と標準偏差（最大値:**太字・下線**, SD: 標準偏差）

性別	ジャンル	手法 1(SD)	手法 2(SD)	手法 3(SD)	手法 4(SD)
男性	アート・建築・デザイン	**2.89** (1.13)	2.65 (1.13)	2.68 (0.95)	**2.89** (1.12)
	コミックス	**3.11** (1.04)	2.54 (1.17)	2.68 (1.17)	3.05 (1.13)
	サイエンス・テクノロジー	**3.06** (0.94)	2.60 (1.12)	2.68 (1.01)	2.95 (1.00)
	スポーツ・アウトドア	**3.08** (1.08)	2.66 (1.22)	2.46 (1.22)	3.00 (1.21)
	ノンフィクション	**2.99** (1.21)	2.61 (1.13)	2.67 (1.16)	2.73 (1.12)
	ビジネス・経済	**3.16** (1.14)	2.84 (1.12)	2.55 (1.19)	3.14 (1.12)
	音楽	**3.39** (1.06)	2.69 (1.04)	2.80 (1.09)	3.19 (1.17)
	絵本・児童書	3.33 (1.08)	2.49 (1.00)	2.72 (1.13)	**3.36** (1.06)
	教育・自己啓発	**3.12** (1.18)	2.80 (1.13)	2.46 (1.01)	2.55 (1.15)
	芸能・エンターテインメント	2.72 (1.10)	2.56 (1.01)	2.60 (1.12)	**2.81** (1.22)
	事典・年鑑・本・ことば	**3.15** (1.03)	2.61 (1.09)	2.88 (1.08)	**3.15** (1.29)
	趣味・実用	**2.99** (1.02)	2.40 (1.08)	2.67 (1.18)	2.75 (1.21)
	人文・思想・宗教	**3.18** (1.05)	2.84 (1.10)	2.92 (1.12)	3.07 (1.18)
	政治・社会	**2.98** (0.93)	2.85 (1.10)	2.81 (1.22)	2.75 (1.17)
	文芸・評論	**2.91** (1.08)	2.64 (1.10)	2.54 (1.12)	2.89 (0.95)
	暮らし・健康・料理	**3.04** (1.09)	2.71 (1.10)	3.02 (1.24)	2.80 (1.03)
	旅行・紀行	2.91 (1.08)	2.79 (1.14)	2.75 (1.08)	**2.99** (1.01)
	歴史・地理	**3.24** (1.02)	2.62 (1.12)	2.66 (1.16)	3.08 (1.17)
	男性平均	**2.98** (1.08)	2.66 (1.11)	2.64 (1.14)	2.91 (1.15)
	最大値の合計	**15**	0	0	5
女性	アート・建築・デザイン	2.78 (1.28)	2.34 (1.08)	2.55 (1.20)	**2.86** (1.29)
	コミックス	**2.95** (1.09)	2.38 (1.25)	2.36 (1.07)	2.91 (1.16)
	サイエンス・テクノロジー	**3.18** (1.04)	2.21 (1.03)	2.54 (1.19)	2.87 (1.21)
	スポーツ・アウトドア	**2.88** (1.15)	2.72 (1.25)	2.34 (1.09)	**2.88** (1.26)
	ノンフィクション	**3.02** (1.28)	2.34 (1.06)	2.44 (1.13)	2.78 (1.30)
	ビジネス・経済	**3.06** (1.17)	2.14 (1.05)	2.42 (1.08)	2.91 (1.12)
	音楽	**3.15** (1.10)	2.32 (0.99)	2.75 (1.15)	3.14 (1.19)
	絵本・児童書	3.20 (1.10)	2.12 (0.99)	2.69 (1.21)	**3.31** (1.15)
	教育・自己啓発	**2.85** (1.16)	2.51 (1.12)	2.38 (1.10)	2.45 (1.05)
	芸能・エンターテインメント	**2.76** (1.20)	2.65 (1.06)	2.22 (1.16)	2.62 (1.15)
	事典・年鑑・本・ことば	2.87 (1.15)	2.76 (1.11)	2.85 (1.17)	**3.11** (1.37)
	趣味・実用	2.72 (1.07)	2.39 (1.03)	2.67 (1.17)	**2.78** (1.24)
	人文・思想・宗教	**2.99** (1.05)	2.82 (1.10)	2.55 (1.15)	**2.99** (1.22)
	政治・社会	**2.95** (1.04)	2.66 (1.14)	2.59 (1.06)	2.66 (1.19)
	文芸・評論	**2.82** (1.13)	2.60 (1.20)	2.15 (1.01)	2.78 (1.20)
	暮らし・健康・料理	**2.84** (1.03)	2.79 (1.25)	2.76 (1.12)	2.75 (1.18)
	旅行・紀行	2.78 (1.10)	2.55 (1.08)	2.52 (1.17)	**2.99** (1.20)
	歴史・地理	**3.01** (1.18)	2.52 (1.14)	2.76 (1.09)	2.91 (1.15)
	女性平均	**3.03** (1.14)	2.49 (1.13)	2.59 (1.15)	2.92 (1.22)
	最大値の合計	**13**	0	0	7

表 7: 手順 4「相対評価」により得られた性別・カテゴリ別の選択手法（最大数:**太字・下線**, %: 割合）

性別	ジャンル	手法 1(%)	手法 2(%)	手法 3(%)	手法 4(%)
男性	アート・建築・デザイン	**29** (34.1%)	25 (29.4%)	11 (12.9%)	20 (23.5%)
	コミックス	**25** (29.4%)	**25** (29.4%)	10 (11.8%)	**25** (29.4%)
	サイエンス・テクノロジー	**29** (34.1%)	20 (23.5%)	17 (20.0%)	19 (22.4%)
	スポーツ・アウトドア	23 (27.1%)	**26** (30.6%)	14 (16.5%)	22 (25.9%)
	ノンフィクション	**30** (35.3%)	14 (16.5%)	24 (28.2%)	17 (20.0%)
	ビジネス・経済	**31** (36.5%)	20 (23.5%)	11 (12.9%)	23 (27.1%)
	音楽	**39** (45.9%)	13 (15.3%)	11 (12.9%)	22 (25.9%)
	絵本・児童書	30 (35.3%)	14 (16.5%)	10 (11.8%)	**31** (36.5%)
	教育・自己啓発	**32** (37.6%)	27 (31.8%)	10 (11.8%)	16 (18.8%)
	芸能・エンターテインメント	**26** (30.6%)	17 (20.0%)	23 (27.1%)	19 (22.4%)
	事典・年鑑・本・ことば	22 (25.9%)	16 (18.8%)	21 (24.7%)	**26** (30.6%)
	趣味・実用	**26** (30.6%)	15 (17.6%)	21 (24.7%)	23 (27.1%)
	人文・思想・宗教	24 (28.2%)	**25** (29.4%)	18 (21.2%)	18 (21.2%)
	政治・社会	23 (27.1%)	23 (27.1%)	**25** (29.4%)	14 (16.5%)
	文芸・評論	21 (24.7%)	**23** (27.1%)	20 (23.5%)	21 (24.7%)
	暮らし・健康・料理	**25** (29.4%)	19 (22.4%)	23 (27.1%)	18 (21.2%)
	旅行・紀行	**25** (29.4%)	18 (21.2%)	21 (24.7%)	21 (24.7%)
	歴史・地理	**27** (31.8%)	20 (23.5%)	20 (23.5%)	18 (21.2%)
	男性合計	**487** (31.8%)	360 (23.5%)	310 (20.3%)	373 (24.4%)
	最大数の合計	**12**	3	2	3
女性	アート・建築・デザイン	**28** (32.9%)	14 (16.5%)	17 (20.0%)	26 (30.6%)
	コミックス	**28** (32.9%)	20 (23.5%)	12 (14.1%)	25 (29.4%)
	サイエンス・テクノロジー	**40** (47.1%)	9 (10.6%)	18 (21.2%)	18 (21.2%)
	スポーツ・アウトドア	21 (24.7%)	**28** (32.9%)	11 (12.9%)	25 (29.4%)
	ノンフィクション	**40** (47.1%)	13 (15.3%)	12 (14.1%)	20 (23.5%)
	ビジネス・経済	**38** (44.7%)	13 (15.3%)	15 (17.6%)	19 (22.4%)
	音楽	28 (32.9%)	6 (7.1%)	20 (23.5%)	**31** (36.5%)
	絵本・児童書	**36** (42.4%)	7 (8.2%)	14 (16.5%)	28 (32.9%)
	教育・自己啓発	**31** (36.5%)	26 (30.6%)	14 (16.5%)	14 (16.5%)
	芸能・エンターテインメント	26 (30.6%)	**31** (36.5%)	14 (16.5%)	14 (16.5%)
	事典・年鑑・本・ことば	26 (30.6%)	10 (11.8%)	18 (21.2%)	**31** (36.5%)
	趣味・実用	**25** (29.4%)	18 (21.2%)	21 (24.7%)	21 (24.7%)
	人文・思想・宗教	**23** (27.1%)	22 (25.9%)	17 (20.0%)	**23** (27.1%)
	政治・社会	**29** (34.1%)	20 (23.5%)	14 (16.5%)	22 (25.9%)
	文芸・評論	**36** (42.4%)	18 (21.2%)	10 (11.8%)	21 (24.7%)
	暮らし・健康・料理	**31** (36.5%)	23 (27.1%)	12 (14.1%)	19 (22.4%)
	旅行・紀行	**31** (36.5%)	14 (16.5%)	12 (14.1%)	28 (32.9%)
	歴史・地理	**36** (42.4%)	17 (20.0%)	15 (17.6%)	17 (20.0%)
	女性合計	**553** (36.1%)	309 (20.2%)	266 (17.4%)	402 (26.3%)
	最大数の合計	**14**	2	0	3

4.3.2 性別による傾向の分析

表 6 は，5 段階評価の平均点と SD を性別とカテゴリでクロス集計し，最終行に性別ごとの全体平均と標準偏差を加えたものである．**表 6** によれば，女性は多くのジャンルで手法 1 を最も高く評価し，全体平均でも**3.03**点（SD 1.14）と最も高かった．男性も全体平均では手法 1 が**2.98**点（SD 1.08）と最も高かった．標準偏差を見ると，男性は手法 1 の SD が 1.08 と最も小さく，評価のばらつきが小さかったのに対し，女性は手法 2 の SD が 1.13 と最も小さかった．

表 7 は，最終選択数を性別とカテゴリでクロス集計し，最終行に性別ごとの合計と割合を加えたものである．全体合計で見ると，女性は手法 1 を選択した割合が 36.1%に達したのに対し，男性は 31.8%であった．男性は「音楽」(39 件) や「ノンフィクション」(30 件) などで手法 1 を最も多く選択した．女性は「サイエンス・テクノロジー」(40 件) や「ノンフィクション」(40 件) で手法 1 を最も多く選択したが,「音楽」(31 件) や「事典・年鑑・本・ことば」(31 件) では手法 4 を最も多く選択した．

5 考察

本章では，主観評価実験の結果に基づき，提案した 4 つのプロンプト構成手法の有効性，評価のばらつき，およびカテゴリ・性別による傾向の違いについて考察する．

実験全体の結果（**表 4** および**表 5** の最終行）では，5 段階評価の平均点と最終選択率の順位が一貫しており，手法 1 が最も高い評価（平均点 3.00 点（**表 4**, 全体平均），選択率 34.0%（**表 5**, 全体合計））を獲得した．これは，書籍説明文中の名詞をすべてプロンプトとして利用すること（手法 1, 全名詞 Positive）で，画像生成 AI に網羅的な情報を与え，結果として書籍の内容を意味的に一貫して反映する画像が生成されたためと推察される．全体平均において，手法 1 は標準偏差 1.12（**表 4**, 全体平均）と最も小さく，評価のばらつきが最も小さかった．これは，情報量の多い手法 1（全名詞 Positive）では生成の方向性が安定し，被験者間で評価のコンセンサスが最も得られやすかったことを示唆している．一方で，手法 4（重み付け）は標準偏差が 1.18（**表 4**, 全体平均）と最もばらつきが大きかった．これは，特定の単語を強調する処理が，画像生成 AI に与えた内容の間の一貫性を保ちにくくなり，評価の振れ幅を大きくした可能性を示唆している．

手法 2（Positive/Negative 分割）と手法 3（上位 10 件）は，全体的に低い評価となった（**表 4** の全体平均で，手法 2 は 2.58 点，手法 3 は 2.61 点）． 手法 1 や手法 4 は Positive プロンプトに平均して約 30 語の名詞が含まれていることに対し，手法 3 は，書籍の複雑な内容をわずか 10 語の名詞のみで表現しようとしたため，情報量が不足し，生成画像が書籍の内容との意味的な一貫性を損ねた可能性がある．手法 2 は，重要度の低い

単語を Negative プロンプトに指定することで，かえって内容のバランスを欠いた画像を生成させ，評価を下げる要因となったと推察される．これら手法 2 と手法 3 の結果は，視覚情報とテキスト情報が意味的に一貫しているほど好意的な評価につながるという Coherence Principle[Mayer 05] の観点から，書籍の内容をスムーズに理解させる上で有効ではなかったことを示唆している．

「ノンフィクション」（評価点 3.01，選択数 70 件），「ビジネス・経済」（評価点 3.11，選択数 69 件），「サイエンス・テクノロジー」（評価点 3.12，選択数 69 件）といったジャンルにおいて，手法 1（全名詞 Positive）が最も高い評価と選択数を獲得する傾向が見られた．これらのジャンルの説明文には，実際に「システム」「構造」「社会」「経済」といった，具体的形状を持たない抽象的な単語や，要素間の関係性を示す単語が多く含まれている．説明文中の名詞をすべて均一な重みで網羅的に用いる手法 1 は，特定の単語を突出させることなく，これらの抽象的な単語と背景や周辺要素を表現する単語を平等に扱う．そのため，書籍が持つ「複合的な文脈」や「世界感」を画像全体で表現するのに適していたと推察される．すなわち，テキスト情報の複雑さと画像の視覚的な情報量が一致することで，Coherence Principle[Mayer 05] に基づき，好意的な評価につながった考えられる．

一方で，「事典・年鑑・本・ことば」（選択数 57 件で最多，評価点 3.13 で 2 位）や「絵本・児童書」（評価点 3.34 で最多，選択数 59 件で 2 位）といったジャンルでは，手法 4（重み付け）も高い評価を得た．これらのジャンルでは，複雑な背景描写よりも，特定の「主題（キャラクターやトピック）」が明確であることが表紙デザインとして求められる傾向にある．手法 4 は重要度の高い単語を強調することで，画像生成 AI に明確な焦点を与え，主題が際立った画像を生成する効果がある．これにより，読者が内容を把握しやすくなる「知覚的流暢性（Processing Fluency）」[Reber 04] が高まり，好評価につながったと考えられる．

以上のことから，全ての情報を網羅する手法 1 は「文脈や世界観」を重視するジャンルに，重要語を強調する手法 4 は「主題の明確さ」を重視するジャンルに適しているという，ジャンルごとの最適なプロンプト構成の違いが示唆された．「スポーツ・アウトドア」カテゴリでは，他のジャンルと異なり手法 2 が最も多く選択された（選択率 31.8%, **表 5**）．これは，重要度の低い情報を Negative プロンプトで排除する操作が，かえって主題（アクションや選手など）を際立たせる効果を生んだ可能性がある．

性別による傾向（**表 6**, **表 7**）にも，平均点と選択率のねじれが表れている．男性は，平均点（2.98 点），選択率（31.8%）において手法 1 を最も支持した．一方，女性は，平均点（3.03 点）と選択率（36.1%）では手法 1 を最も支持したが，手法 4 に対しても一部のジャンル（事典・年鑑・本・ことば，音楽など）で強い支持を示した．これは，女性被験者において，手法 4（重み付け）が生成する「焦点が明確な画像」が好まれるケースがあったことを示唆している．

6 結論

本研究では，画像生成 AI を用いて書籍の説明文から表紙画像を生成し，どのプロンプト構成が**書籍表紙の誘引力**を高めるかについて主観評価実験を行った．その結果，説明文中の名詞をすべて用いる手法（手法 1）が全体的に最も好まれる傾向がある一方で，重要語を重み付けする手法（手法 4）も特定のジャンルや性別においては有効であることが明らかになった．この結果は，情報量が多く意味的に一貫した画像や，焦点が明確な画像が好まれやすいという認知心理学の理論によって支持される [Reber 04, Mayer 05]．

今後は，被験者層の拡大や，生成された画像自体の定量的分析に取り組むことで，AI による書籍デザイン支援の可能性をさらに追求できると期待する．

参考文献

[Orth 08] Orth, U. R., & Malkewitz, K.: Holistic package design and consumer brand impressions, *Journal of Marketing*, Vol. 72, No. 3, pp. 64-81 (2008).

[Liu 17] Liu, Y., Li, K., Hu, H.: The impact of positive imagery in visual marketing on consumer approach behavior, *Journal of Business Research*, Vol. 79, pp. 106-114 (2017).

[Bloch 95] Bloch, P. H.: Seeking the ideal form: Product design and consumer response, *Journal of marketing*, Vol. 59, No. 3, pp. 16-29 (1995).

[Bruce 97] Bruce, M., & Cooper, R.: Marketing and design in the new product development process, *Marketing intelligence & planning*, Vol. 15, No. 3, pp. 100-106 (1997).

[Gudinavičius 17] Gudinavičius, A., & Šuminas, A.: Choosing a book by its cover: analysis of a reader's choice, *Journal of Documentation*, Vol. 74, No. 2 (2017).

[Hagtvedt 08] Hagtvedt, H., & Patrick, V. M.: Art and the brand: The role of visual art in enhancing brand extendibility, *Journal of Consumer Psychology*, Vol. 18, No. 3, pp. 212-222 (2008).

[Jian 19] Jian, W., Lyu, S., & Li, Y.: The influence of book cover design on reader's perception of content and quality, *Publishing Research Quarterly*, Vol. 35, pp. 466-476 (2019).

[Jiang 13] Jiang, Z., Chan, J., Tan, B. C., & Chua, W. Y.: Effects of interactivity on website involvement and purchase intention, *Journal of the Association for Information Systems*, Vol. 14, No. 1, p. 3 (2013).

[Mayer 05] Mayer, R. E.: The Cambridge handbook of multimedia learning, Cambridge university press (2005).

[Park 23] Park, J. Y., Kim, C., Park, S., Dio, K.: Do you judge a a book by its cover? Online book purchases between Japan and France, *Asia Pacific Journal of Marketing and Logistics*, Vol. 35, No. 2 (2023).

[Ramesh 21] Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... Sutskever, I.: Zero-shot text-to-image generation, *International Conference on Machine Learning*, PMLR, pp. 8821-8831 (2021).

[Reber 04] Reber, R., Schwarz, N., Winkielman, P.: Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?, *Personality and social psychology review*, Vol. 8, No. 4, pp. 364-382 (2004).

[Rombach 22] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684-10695 (2022).

[Silayoi 07] Silayoi, P., & Speece, M.: The importance of packaging attributes: a conjoint analysis approach, *European Journal of Marketing*, Vol. 41, No. 11/12, pp. 1495-1517 (2007).