

geDIG: 自由エネルギー原理に基づく RAG の適応的検索制御

geDIG: Adaptive Retrieval Control for RAG Based on Free Energy Principle

宮内 和義 ^{*1}

Kazuyoshi Miyauchi

^{*1}独立研究者

Independent Researcher

Retrieval-Augmented Generation (RAG) lacks principled criteria for deciding “when” to perform additional retrieval. We propose **geDIG**, a unified gauge bridging the Free Energy Principle (FEP) and Minimum Description Length (MDL), defined as $\mathcal{F} = \Delta EPC_{norm} - \lambda \cdot \Delta IG$. This gauge enables autonomous control via two gates: the Attention Gate (AG) triggers exploration upon detecting ambiguity, while the Decision Gate (DG) confirms integration when structural shortcuts emerge. On the HotpotQA benchmark (28,954 questions), geDIG achieves Exact Match 37.1% and F1 53.4%, outperforming the BM25 baseline by +3.5pt EM and +4.0pt F1. Furthermore, validation on partial-observation maze tasks demonstrates that the AG/DG mechanism functions independently of language, achieving 98% success rate on 15×15 mazes. geDIG provides a principled “When” criterion for knowledge graph updates in RAG systems.

1. はじめに

Retrieval-Augmented Generation (RAG) [4] は、大規模言語モデル (LLM) の知識限界を外部検索で補完する手法として広く普及している。しかし、現行の RAG システムには根本的な課題が残されている：「いつ追加検索すべきか」の判断基準が存在しない。

多くの RAG システムは固定回数の検索を行うか、LLM の自己判断に依存している。このアドホックなアプローチは、過少検索（必要な情報の欠落）、過剰検索（不要なコスト）、判断の不透明性という問題を引き起こす。特にマルチホップ推論を要する質問（例：HotpotQA[3]）では、何回検索すれば十分かの判断が困難である。

本研究では、自由エネルギー原理 (FEP) [1] と最小記述長 (MDL) [2] を操作的に橋渡しする統一ゲージ **geDIG** (Graph Edit Distance + Information Gain) を提案する。geDIG は知識グラフの構造変化を単一スカラーで評価し、「探索すべきか」「統合すべきか」の判断を自律的に行う。

本稿の貢献は以下の 3 点である：

1. FEP-MDL を橋渡しする統一ゲージ geDIG の提案
2. HotpotQA ベンチマークでの有効性実証 (EM +3.5pt, F1 +4.0pt vs BM25)
3. 部分観測迷路での言語非依存な原理検証 (成功率 98%)

2. 提案手法：geDIG

2.1 統一ゲージの定義

geDIG は、知識グラフの更新前後での構造変化と情報利得をトレードオフする：

$$\mathcal{F} = \Delta EPC_{norm} - \lambda \cdot \Delta IG \quad (1)$$

連絡先：宮内 和義（独立研究者），E-mail:
miyauchi.kazuyoshi@gmail.com

表 1: geDIG 構成項の理論的対応

項	意味	FEP/MDL 対応
ΔEPC_{norm}	構造編集コスト	モデル記述長 $L(M)$
ΔH	エントロピー変化	変分自由エネルギー
ΔSP	最短路ゲイン	データ記述長 $L(D M)$

表 2: 従来手法との比較：評価対象

手法	確率（類似度）	構造（グラフ）
BM25 / Contriever	✓	—
GraphRAG	—	✓
geDIG	✓	✓

ここで ΔEPC_{norm} は正規化グラフ編集距離（ノード・エッジの追加コスト）， ΔIG は情報利得（エントロピー減少+最短路ゲイン）， λ はトレードオフパラメータである。

解説： \mathcal{F} が小さいほど「コストに見合った情報整理」が達成されている。

2.2 なぜ構造と確率の同時評価が必要か

従来手法は、類似度（確率的指標）またはグラフ構造のいずれか一方のみを評価する（表 2）。BM25 や Contriever は文書間の類似度スコアのみで検索を打ち切るため、「高スコアだが冗長な情報」を排除できない。一方、GraphRAG はグラフ構造を活用するが、「構造は整っているが不確実性が残る」状態を検知できない。

geDIG の核心は、構造変化のコスト (ΔEPC_{norm}) と確率的な情報利得 (ΔIG) の収支を単一スカラーで評価する点にある。直感的には、「新しいエッジを張る構造的コストに見合うだけの不確実性削減が得られるか？」という問い合わせに答える。この収支計算により、以下の判断が可能になる：

- 構造コスト大・情報利得小 → 検索不要（冗長）
- 構造コスト小・情報利得大 → 検索有効（統合）
- 構造コスト大・情報利得大 → 探索継続（AG 発火）

```

[Query] → AG 評価
  ↪  $g_0 > \theta_{AG}$ ? → Yes: 追加検索
  ↪ No: DG 評価
    ↪  $g_{min} \leq \theta_{DG}$ ? → Yes: 回答生成
    ↪ No: 追加検索

```

図 1: AG/DG 制御フロー

表 3: HotpotQA 結果

手法	サンプル数	EM	F1
Closed-book	5,511	22.7%	35.5%
Contriever	599	25.5%	40.4%
BM25	1,110	33.6%	49.4%
geDIG	28,954	37.1%	53.4%
Static GraphRAG	600	40.3%	57.4%

この「いつ」の判断は、構造と確率を同時に見なければ下せない。これが geDIG の本質的な新規性である。

2.3 AG/DG 二段ゲート

geDIG に基づき、2種類のゲートで検索制御を行う（図 1）。
AG (Attention Gate) : 0-hop 評価で曖昧さを検知する。
 $g_0 > \theta_{AG}$ のとき、追加検索を実行する。これは FEP における「予測誤差」の検出に対応する。

DG (Decision Gate) : Multi-hop 評価で安定性を確認する。 $g_{min} \leq \theta_{DG}$ のとき、回答生成を実行する。これは MDL における「記述長削減の収束」に対応する。

3. 実験 I : HotpotQA ベンチマーク

3.1 実験設定

データセット: HotpotQA Distractor Dev Set[3] (7,405 問) を使用した。各問題に付属する context のみを検索対象とする閉世界設定で評価した。

比較手法 :

- **Closed-book** : 検索なし、LLM の知識のみ
- **BM25** : 固定 Top-5 検索
- **Contriever[5]** : 密ベクトル検索 (Top-5)
- **Static GraphRAG** : 事前構築グラフからの検索
- **geDIG** : AG/DG による適応的検索

全手法で GPT-4o-mini を回答生成に使用した。geDIG のパラメータは $\lambda = 0.3$, $\theta_{AG} = 0.5$, $\theta_{DG} = 0.3$, $\text{max_hops} = 3$ とした。

3.2 結果

表 3 に主要結果を示す。geDIG は BM25 ベースラインに対して、Exact Match (EM) で +3.5pt, F1 で +4.0pt の改善を達成した。

表 4 に geDIG のゲート動作統計を示す。AG 発火率 45.6% は、約半数の問題で追加検索が必要と判断されたことを示す。平均グラフエッジ数 10.0 は、過剰なグラフ拡張が抑制されていることを示唆する。

表 4: geDIG ゲート動作統計

指標	値
AG 発火率 (初期)	45.6%
DG 発火率 (初期)	36.5%
AG 発火率 (最終)	43.9%
DG 発火率 (最終)	25.9%
平均 geDIG スコア	0.489
平均グラフエッジ数	10.0

表 5: 迷路サイズ別結果

サイズ	シード数	成功率	平均エッジ
15×15	100	98%	69.0
25×25	60	75%	284.5
51×51	13	46%	1541.0

4. 実験 II : 迷路による原理検証

geDIG のゲート機構が言語に依存せず機能することを確認するため、部分観測迷路での検証を行った。エージェントは周囲 1 マスのみ観測可能であり、移動しながら内部グラフを構築する。

表 5 に結果を示す。15×15 迷路で成功率 98% を達成し、AG/DG の二段制御が探索と統合を適切に切り替えることを確認した。51×51 ではステップ上限の影響で成功率が低下したが、これは探索予算の制約によるものである。

5. 考察

5.1 geDIG の有効性

HotpotQA での改善 (EM +3.5pt, F1 +4.0pt) は、適応的検索制御が固定回数検索より効果的であることを示す。AG 発火率 45.6% は、全問題の約半数で追加検索が有効と判断されたことを意味し、残り半数では初回検索で十分であったことを示唆する。

5.2 Static GraphRAG との比較

Static GraphRAG が geDIG を上回る結果 (EM 40.3% vs 37.1%) は、事前構築された密なグラフ構造の有効性を示す。一方、geDIG は動的にグラフを構築するため計算効率が高い。両者の組み合わせ (事前構築+動的拡張) が今後の方向性として考えられる。

5.3 限界と今後の課題

1. 閾値 θ_{AG} , θ_{DG} の自動調整手法
2. オフライン再配線 (Phase 2) の実装
3. 大規模 LLM (GPT-4, Claude 等) での検証

6. 関連研究

RAG の検索回数制御に関しては、Self-RAG[7] が LLM の自己判断による適応的検索を提案している。本研究は LLM 非依存の原理的基準を提供する点で異なる。Graph RAG[6] は知識グラフを用いた検索拡張を提案しているが、動的なグラフ更新の判断基準は示していない。geDIG はこの「When」問題に対する原理的な解を提供する。

7. おわりに

本研究では、RAGにおける「いつ検索すべきか」の問題に対し、自由エネルギー原理に基づく統一ゲージ geDIGを提案した。HotpotQA ベンチマーク(28,954問)において、geDIGはBM25ベースラインに対してEM +3.5pt, F1 +4.0ptの改善を達成した。また、部分観測迷路での検証により、AG/DG二段ゲートが言語非依存で機能することを確認した。geDIGは、知識グラフの「When」問題に対する原理的な解を提供し、RAGの自律的制御への基盤となる。

参考文献

- [1] Friston, K.: The free-energy principle: a unified brain theory?, Nat. Rev. Neurosci., 11, 127–138 (2010).
- [2] Grünwald, P. D.: The Minimum Description Length Principle, MIT Press (2007).
- [3] Yang, Z., et al.: HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering, EMNLP (2018).
- [4] Lewis, P., et al.: Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, NeurIPS (2020).
- [5] Izacard, G., et al.: Unsupervised Dense Information Retrieval with Contrastive Learning, TMLR (2022).
- [6] Edge, D., et al.: From Local to Global: A Graph RAG Approach, arXiv (2024).
- [7] Asai, A., et al.: Self-RAG: Learning to Retrieve, Generate, and Critique, NeurIPS (2023).