

マルチモーダル知識ハイパーグラフを利用した 生物医学分野における知識拡張情報抽出

西出隆盛 三輪誠
豊田工業大学大学院

{sd25428,makoto-miwa}@toyota-ti.ac.jp

概要

生物医学分野では化学構造やタンパク質配列といったマルチモーダルな分野知識が知識ベースとして整備されている。しかし、既存の情報抽出における知識拡張手法はテキスト情報を中心としており、このような知識を十分に活用できていない。また、知識は互いに複雑に関係しており、単純なグラフ構造では表現が困難である。本研究では、知識ベースからマルチモーダル知識ハイパーグラフを構築し、ハイパーグラフニューラルネットワークで表現学習を行い、その表現を言語モデルに統合する手法を提案する。実験では、生物医学情報抽出ベンチマークにおいてマルチモーダル知識ハイパーグラフの有効性を確認した。

1 はじめに

生物医学分野において、文献から構造化された知識を獲得する情報抽出は、膨大な文献に含まれる情報を整理するための重要な技術として長らく研究されてきた。固有表現抽出 (Named Entity Recognition; NER) や関係抽出 (Relation Extraction; RE) などの技術を用いて自動的に抽出された遺伝子、タンパク質、疾患、薬剤といったエンティティやその関係は、知識ベースの構築や仮説生成の支援に役立つと期待されている。

近年、事前学習済み言語モデルや大規模言語モデルを用いた情報抽出が発展している。しかし、生物医学分野の文書の解析には事前学習のテキストには含まれない外部の様々な分野知識が必要である。例えば、生物医学分野では、同一のエンティティが文献中で異なる表記で言及されることが多く (例: 「aspirin」「acetylsalicylic acid」), その正規化には標準化のための知識ベースが必要となる。また、化学物質の異性体のように、テキストのみでは区別が困難

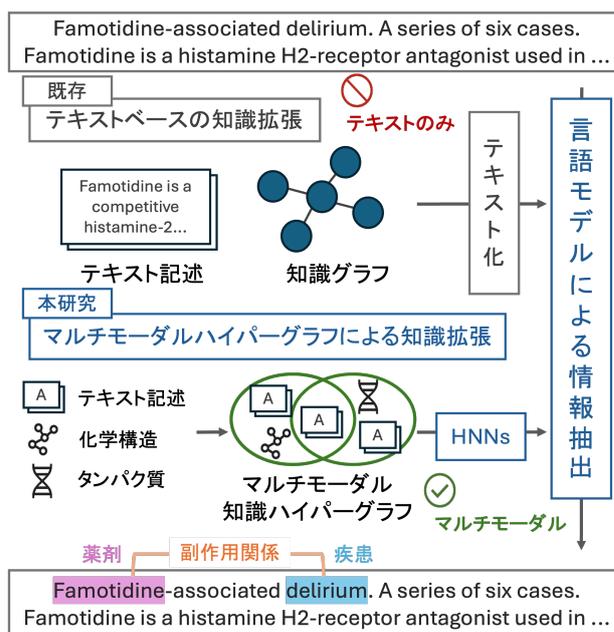


図1 既存手法と提案手法の比較。既存の知識拡張はテキスト化された情報のみを使用するが、提案手法はマルチモーダル知識ハイパーグラフにより化学構造やタンパク質情報を直接活用する。

な概念も存在し、化学構造などの情報が必要となる。さらに、希少疾患や新規化合物が記述された文書は限られるため、外部知識による補完が重要である。

言語モデルの推論における知識拡張手法として、外部の知識ベースを推論時に参照する検索拡張手法が提案されている。例えば、GraphRAG [1, 2] は知識グラフにコミュニティ検出を行い、構造的な関係性を活用して推論経路を言語モデルに提供する。HyperGraphRAG [3] は、ハイパーグラフを用いて任意個の要素間の関係を保持したまま情報を検索することで、複雑な関係の理解を可能にする。

一方で、既存の検索拡張手法はテキスト情報を中心としており、生物医学分野の外部知識に含まれるマルチモーダル情報を十分に活用できていない。生

物医学分野の外部知識には、テキスト記述に加えて、化学構造式 (SMILES) やタンパク質配列といったマルチモーダル情報を含む知識が互いに関係している。例えば、「薬剤 A と薬剤 B の併用が特定の遺伝子型を持つ患者において副作用 C を引き起こす」といった関係では、こうしたマルチモーダル情報が関係の理解に不可欠である。

本研究では、このような互いに複雑に関係したマルチモーダルな外部知識の言語モデルでの活用を目的に、マルチモーダル知識ハイパーグラフを用いた情報抽出手法を提案する。マルチモーダル知識グラフを基盤としてハイパーグラフを構築し、任意個の要素間のまとまりをハイパーエッジとして表現する。具体的には、ハイパーグラフニューラルネットワーク (HNNs) [4] で表現を学習し、言語モデルに統合することで情報抽出を実現する。

本研究の貢献は以下の通りである。

- マルチモーダル属性を持つ知識グラフから、複雑な関係を保持するハイパーグラフを構築する手法を示した。
- ハイパーグラフニューラルネットワークと言語モデルを統合するアーキテクチャを設計した。
- 生物医学情報抽出ベンチマークにおいて、提案手法の有効性を検証した。

2 関連研究

2.1 外部知識を活用した情報抽出

生物医学 NER において、外部知識を動的に活用する手法が提案されている。kNN-NER [5] は類似事例の検索により希少エンティティの認識を向上させ、REAL [6] は LLM と生物医学オントロジーを組み合わせた検索拡張を適用した。これらの手法はテキスト情報を外部知識としており、化学構造や配列情報の活用には至っていない。

2.2 知識グラフと検索拡張手法

GraphRAG [1, 2] は、知識グラフにコミュニティ検出を適用し、LLM で要約を生成することで、大域的な推論を可能にした。HippoRAG [7, 8] は、Personalized PageRank によりグラフ上の連想的探索を行った。HyperGraphRAG [3] は、ハイパーグラフを用いて N 項関係を保持したまま、生成タスクにおける文脈を提供する手法を提案した。これらの手法

においても、知識グラフに含まれる化学構造や配列といったマルチモーダル情報は活用できていない。

2.3 マルチモーダル知識グラフ

PrimeKG [9] は、20 の知識ベースを統合した 129,375 ノード、400 万エッジの大規模知識グラフである。PrimeKG++ [10] は、さらに、各ノードに SMILES、アミノ酸配列、テキスト記述を付与し、各モダリティデータに対し特化型エンコーダで処理して得た表現を用いた、グラフ対照学習により、テキストに明示されていない暗黙的な関係を発見した。

マルチモーダル知識グラフの表現学習では、異なるモダリティ間の融合が重要となる。NativE [11] は、関係に応じてモダリティの重みを動的に調整する ReDAF (Relation-guided Dual Adaptive Fusion) [11] を提案した。浅田ら [12] は薬物の分子構造情報と知識グラフを DDI 抽出に活用したが、グラフ内の高次の関係は考慮していない。

2.4 ハイパーグラフ表現学習

ハイパーグラフは任意の数のノードを結ぶハイパーエッジを用いて複雑な高次関係を表現するデータ構造である。ハイパーグラフニューラルネットワーク (HNNs) は、ハイパーグラフ上でメッセージパッシングを実現し、ノードとハイパーエッジの表現を学習する。SE-HSSL [13] は、構造学習とハイパーグラフ自己教師あり学習を組み合わせ、グラフの構造的パターンを学習した。

3 提案手法

本研究では、PrimeKG++ [10] を基盤としたマルチモーダル知識ハイパーグラフを構築し、事前に自己教師あり学習による学習を行なった HNNs を通じて得た表現を言語モデルに統合することでハイパーグラフの知識を利用した推論を行う。

3.1 マルチモーダル知識ハイパーグラフの構築

本研究では、PrimeKG++ [10] を基盤として、生物学的な N 項関係を表現するマルチモーダル知識ハイパーグラフを構築する。PrimeKG++ は 129,312 個のエンティティと約 810 万の二項関係から構成され、各エンティティには SMILES (化学構造)、アミノ酸配列、テキスト記述といったマルチモーダル属性が付与されている。ハイパーエッジの生成は 2 階

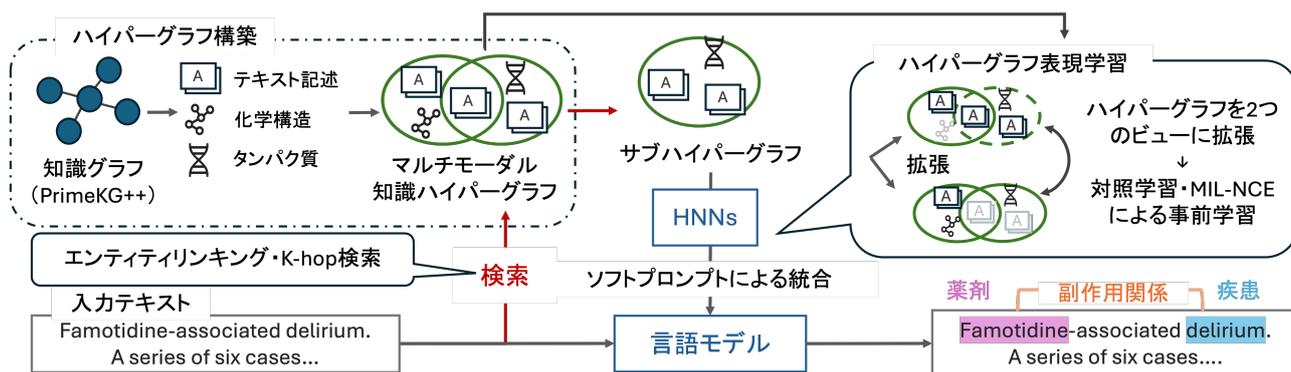


図2 提案手法の概要. 知識グラフからマルチモーダル知識ハイパーグラフを構築し, 入力テキストに対応するサブハイパーグラフを抽出する. HNNs による事前学習で表現を学習し, ソフトプロンプトにより言語モデルに統合する.

層で行う. 1階層目では, 各エンティティとそれを表すマルチモーダルなノード (化学構造ノード, 配列ノード) と結ぶ定義型ハイパーエッジを生成する. 2階層目では, 二項関係を意味論的に統合し, 作用機序 (薬剤-タンパク質-疾患), 中心性ハブ (多数の関係を持つ薬剤・疾患・曝露因子), 生物学的パスウェイ, オントロジー階層, 化学構造の類似性によるクラスタ, タンパク質相互作用の6つのパターンでハイパーエッジを構築する. 各モダリティは特化型エンコーダで埋め込みを取得し, 共通次元に射影する (詳細は付録A).

3.2 ハイパーグラフ表現学習

各ノードのマルチモーダル属性 (テキスト, 化学構造, 配列) は, ReDAF [11] に基づく融合機構により, 関係タイプに応じて適応的に重み付けし, 統合する (詳細は付録B). 融合後のノード表現に対して, SE-HSSL [13] を基盤としたハイパーグラフニューラルネットワーク (HNNs) を適用する. HNNs は2段階のメッセージパッシングでノード表現を学習する. まず, ハイパーエッジ内のノード表現を生成し, 次に各ノードが所属するハイパーエッジの表現を統合してしてノード表現を更新する.

事前学習では, 特徴ドロップと接続ドロップによるデータ拡張を行い, ノードレベルとグループレベルの対照学習損失および, ノードとハイパーエッジ間のメンバーシップを予測する MIL-NCE 損失により, 高次構造を考慮した表現を獲得する (詳細は付録B).

3.3 言語モデルへの統合

学習済みの HNNs を言語モデルに統合し, 入力テキストに関連するサブハイパーグラフの知識を活用

した情報抽出を実現する.

3.3.1 パターンベースのサブハイパーグラフ抽出

入力テキストから軽量の NER モデル [14] でエンティティメンションを抽出し, SapBERT [15] によりハイパーグラフのエンティティにリンクする. リンクされたエンティティ E_{seed} を起点として, 関連するハイパーエッジを選択しサブハイパーグラフ H_{sub} を構築する. まず, E_{seed} に接続する全てのハイパーエッジを候補として収集する. 次に, 各ハイパーエッジにパターンタイプに応じた関連度スコアを付与する (詳細は付録A). 作用機序やパスウェイなどの高次関係を優先し, エンティティ定義のような単純な関係の優先度を下げることで, 生物学的な関係に関わる構造を優先的に抽出する. 最後に, スコア上位 N 個のハイパーエッジを選択してサブハイパーグラフを構成する.

3.3.2 ソフトプロンプトによる知識注入

抽出したサブハイパーグラフを HNNs でエンコードし, ノード表現とハイパーエッジ表現の両方をプーリングしてグラフ埋め込み \mathbf{g} を得る.

$$\mathbf{g} = \frac{1}{2} \left(\frac{1}{|V|} \sum_{v \in V} \mathbf{h}_v + \frac{1}{|E|} \sum_{e \in E} \mathbf{h}_e \right) \quad (1)$$

この埋め込みを線形射影により言語モデルの埋め込みの次元 d に変換し, K 個のソフトプロンプトトークン $\mathbf{P} \in \mathbb{R}^{K \times d}$ を生成する.

$$\mathbf{P} = \text{MLP}(\mathbf{g}), \quad \mathbf{P} \in \mathbb{R}^{K \times d} \quad (2)$$

生成されたソフトプロンプトを入力テキストの埋め込みの先頭に結合し, 言語モデルに入力する.

$$\mathbf{H}_{input} = [\mathbf{P}; \mathbf{E}_{text}] \quad (3)$$

表1 BC5CDR NER タスクの結果

手法	P	R	F1 (%)
Base (ModernBERT)	86.96	86.96	86.87
+ Text Retrieval	86.58	88.11	87.34
+ Graph Text	87.78	87.64	87.71
+ HG Text	87.54	87.89	87.72
Proposed	87.32	88.33	87.82
Proposed + HG Text	88.06	87.24	87.65

表2 BC5CDR RE タスク (CID 関係抽出) の結果

手法	P	R	F1 (%)
Base (ModernBERT)	68.85	68.99	68.92
+ Text Retrieval	70.12	70.12	68.85
+ Graph Text	68.35	69.81	69.07
+ HG Text	69.32	67.82	68.56
Proposed	73.91	70.71	72.28
Proposed + HG Text	71.21	70.65	70.93

これにより、グラフから抽出されたマルチモーダル知識を言語モデルの文脈として機能させる。

4 実験

提案手法の有効性を検証するため、生物医学情報抽出ベンチマーク BC5CDR [16] で評価を行う。BC5CDR は化学物質と疾患の NER タスクおよび化学物質誘発疾患 (CID) の関係抽出タスクを含む。言語モデルには ModernBERT base [17] を使用し、Precision, Recall, F1 スコアで評価する。比較手法として、ベースラインである ModernBERT 単体に加え、検索拡張手法として Qwen3-Embedding [18] によるテキスト検索 (Text Retrieval)、パターンベースの K-hop グラフ検索 (Graph Text)、ハイパーグラフ検索 (HG Text) を用意した。これらはすべて取得したテキストをコンテキスト (詳細は付録 C) として言語モデルに入力する。サブグラフ、およびサブハイパーグラフについては含まれるエンティティ名、定義、関係をテキスト化して利用する。提案手法では、HNNs の出力をソフトプロンプトで統合する手法 (Proposed) と、さらに HG Text を追加する手法 (Proposed + HG Text) を評価する。

表 1 と表 2 に NER および RE タスクの結果を示す。提案手法は両タスクにおいてベースラインを上回り、特に RE タスクにおいて改善が大きくなった。

4.1 アブレーション

提案手法の各コンポーネントの寄与を分析する。表 3 にハイパーグラフ事前学習とソフトプロンプト統合による性能への影響を示す。ソフトプロンプト有りの場合は 8 トークン、ない場合は 1 トークンに

表3 アブレーション結果 (BC5CDR, F1 (%))

Pre-train	Soft Prompt	NER	RE
✓	✓	87.82	72.28
✓		87.16	70.18
	✓	87.69	71.59
		87.10	70.82

射影して入力に結合する。

全コンポーネントを含む提案手法が最も高い性能を示した。ソフトプロンプト統合は、複数トークンの分散表現により線形射影よりも高い表現力が得られたと考えられる。ハイパーグラフ事前学習の有無については性能の向上が限定的であった。これは事前学習がハイパーグラフ全体を扱うのに対して、情報抽出タスクにおける入力がサブハイパーグラフであったことによる分布シフトが一因と考えられる。

4.2 解析

NER において、提案手法はベースラインに比べ Recall を 1.15%ポイント改善した。ハイパーグラフの作用機序やパスウェイ情報により、ベースラインでは見逃されていたエンティティが認識されるようになったと考えられる。RE では、提案手法がベースラインを F1 スコアで 3.36%ポイントの改善を示した。この改善には、マルチモーダル知識の効果があると考えられる。光学異性体のようなテキストのみでは区別できない化合物も、SMILES 埋め込みにより区別可能となる。一方、Proposed + HG Text の性能低下は、テキスト情報の追加がノイズとなったことを示唆している。事例分析は付録 D を参照。

5 おわりに

本研究では、マルチモーダル知識ハイパーグラフを用いた生物医学情報抽出手法を提案した。知識グラフから高次関係をハイパーエッジとして表現し、SMILES 埋め込みやタンパク質配列埋め込みを統合したハイパーグラフを構築した。HNNs による表現学習とソフトプロンプトを用いた言語モデルへの統合により、BC5CDR ベンチマークにおいて、NER で 0.95%ポイント、RE で 3.36%ポイントの F1 スコアの改善を達成した。実験を通じて、テキストのみでは区別困難な光学異性体において、マルチモーダル埋め込みが有効に機能することが示唆された。一方で、現在の手法はベースラインの知識抽出手法の性能とハイパーグラフの検索性能に依存しており、これらの改善が今後の課題である。

謝辞

本研究の一部は、キオクシア株式会社の支援をうけて実施したものです。

参考文献

- [1] Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, Dasha Metropolitan, Robert Osazuwa Ness, and Jonathan Larson. From local to global: A graph rag approach to query-focused summarization, 2025.
- [2] Boci Peng, Yun Zhu, Yongchao Liu, Xiaohe Bo, Haizhou Shi, Chuntao Hong, Yan Zhang, and Siliang Tang. Graph retrieval-augmented generation: A survey, 2024.
- [3] Haoran Luo, Haihong E, Guanting Chen, Yandan Zheng, Xiaobao Wu, Yikai Guo, Qika Lin, Yu Feng, Zemin Kuang, Meina Song, Yifan Zhu, and Luu Anh Tuan. Hypergraphrag: Retrieval-augmented generation via hypergraph-structured knowledge representation, 2025.
- [4] Sunwoo Kim, Soo Yong Lee, Yue Gao, Alessia Antelmi, Mirko Polato, and Kijung Shin. A survey on hypergraph neural networks: An in-depth and step-by-step guide. In **Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '24**, p. 6534–6544. ACM, August 2024.
- [5] Shuhe Wang, Xiaoya Li, Yuxian Meng, Tianwei Zhang, Rongbin Ouyang, Jiwei Li, and Guoyin Wang. *knn-ner*: Named entity recognition with nearest neighbor search, 2022.
- [6] Darya Shlyk, Tudor Groza, Marco Mesiti, Stefano Montanelli, and Emanuele Cavalleri. REAL: A retrieval-augmented entity linking approach for biomedical concept recognition. In Dina Demner-Fushman, Sophia Ananiadou, Makoto Miwa, Kirk Roberts, and Junichi Tsujii, editors, **Proceedings of the 23rd Workshop on Biomedical Natural Language Processing**, pp. 380–389, Bangkok, Thailand, August 2024. Association for Computational Linguistics.
- [7] Bernal Jiménez Gutiérrez, Yiheng Shu, Yu Gu, Michihiro Yasunaga, and Yu Su. Hipporag: Neurobiologically inspired long-term memory for large language models. In **The Thirty-eighth Annual Conference on Neural Information Processing Systems**, 2024.
- [8] Bernal Jiménez Gutiérrez, Yiheng Shu, Weijian Qi, Sizhe Zhou, and Yu Su. From rag to memory: Non-parametric continual learning for large language models, 2025.
- [9] Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. **Nature Scientific Data**, 2023.
- [10] Tien Dang, Viet Thanh Duy Nguyen, Minh Tuan Le, and Truong-Son Hy. Multimodal contrastive representation learning in augmented biomedical knowledge graphs, 2025.
- [11] Yichi Zhang, Zhuo Chen, Lingbing Guo, Yajing Xu, Binbin Hu, Ziqi Liu, Wen Zhang, and Huajun Chen. Native: Multi-modal knowledge graph completion in the wild. In **SIGIR**, pp. 91–101. ACM, 2024.
- [12] Masaki Asada, Makoto Miwa, and Yutaka Sasaki. Integrating heterogeneous knowledge graphs into drug–drug interaction extraction from the literature. **Bioinformatics**, Vol. 39, No. 1, p. btac754, 11 2022.
- [13] Fan Li, Xiaoyang Wang, Dawei Cheng, Wenjie Zhang, Ying Zhang, and Xuemin Lin. Hypergraph self-supervised learning with sampling-efficient signals. **IJCAI '24**, 2024.
- [14] Shaina Raza, Deepak John Reji, Femi Shajan, and Syed Raza Bashir. Large-scale application of named entity recognition to biomedicine and epidemiology. **PLOS Digital Health**, Vol. 1, No. 12, p. e0000152, 2022.
- [15] Fangyu Liu, Ehsan Shareghi, Zaiqiao Meng, Marco Basaldella, and Nigel Collier. Self-alignment pretraining for biomedical entity representations. In **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 4228–4238, Online, June 2021. Association for Computational Linguistics.
- [16] Jiao Li, Yueping Sun, Robin J. Johnson, Daniela Sciaky, Chih-Hsuan Wei, Robert Leaman, Allan Peter Davis, Carolyn J. Mattingly, Thomas C. Wieggers, and Zhiyong Lu. Biocreative V CDR task corpus: a resource for chemical disease relation extraction. **Database J. Biol. Databases Curation**, Vol. 2016, , 2016.
- [17] Benjamin Warner, Antoine Chaffin, Benjamin Clavié, Orion Weller, Oskar Hallström, Said Taghadouini, Alexis Gallagher, Raja Biswas, Faisal Ladhak, Tom Aarsen, Nathan Cooper, Griffin Adams, Jeremy Howard, and Iacopo Poli. Smarter, better, faster, longer: A modern bidirectional encoder for fast, memory efficient, and long context finetuning and inference, 2024.
- [18] Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, Fei Huang, and Jingren Zhou. Qwen3 embedding: Advancing text embedding and reranking through foundation models. **arXiv preprint arXiv:2506.05176**, 2025.
- [19] Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. Domain-specific language model pre-training for biomedical natural language processing. **ACM Trans. Comput. Healthcare**, Vol. 3, No. 1, October 2021.
- [20] Jerret Ross, Brian Belgodere, Vijil Chenthamarakshan, Inkit Padhi, Youssef Mroueh, and Payel Das. Large-scale chemical language representations capture molecular structure and properties. **Nature Machine Intelligence**, Vol. 4, No. 12, pp. 1256–1264, 2022.
- [21] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Sal Candido, et al. Language models of protein sequences at the scale of evolution enable accurate structure prediction. **bioRxiv**, 2022.

A ハイパーグラフ統計

表 4 に構築したハイパーグラフのハイパーエッジパターン別統計を示す。スコアはサブハイパーグラフ抽出時の優先度を表し、作用機序やパスウェイなど生物学的関係を含むパターンを高く設定している。テキストには PubMedBERT[19], SMILES には MolFormer[20], 配列には ESM-2[21] を使用してエンコードする。

B 表現学習の詳細

ReDAF [11] に基づく融合では、共通の隠れ次元 (実験では $d = 512$) に射影した後の各モダリティ埋め込み $\mathbf{e}_{\text{text}}, \mathbf{e}_{\text{mol}}, \mathbf{e}_{\text{seq}}$ を関係タイプ r に応じて重み付けする。

$$\mathbf{e}_{\text{fused}} = \sum_{m \in \mathcal{M}} \alpha_m(r) \cdot \mathbf{e}_m \quad (4)$$

ここで、 $\alpha_m(r)$ は関係タイプ r に応じたモダリティ m の重みである。

SE-HSSL による自己教師あり学習では、ノードレベル CCA 損失, グループレベル CCA 損失, MIL-NCE 損失を用いる。MIL-NCE 損失は以下の式で定義される。

$$\mathcal{L}_{\text{MIL}} = -\log \frac{\sum_{e \in \mathcal{E}_v^+} \exp(\mathbf{z}_v^\top \mathbf{z}_e / \tau)}{\sum_{e \in \mathcal{E}} \exp(\mathbf{z}_v^\top \mathbf{z}_e / \tau)} \quad (5)$$

ここで、 \mathcal{E}_v^+ はノード v が所属するハイパーエッジの集合である。

C 検索コンテキストの例

表 5 に各検索手法のコンテキストの例を示す。Text Retrieval は入力テキストと意味的に類似したエンティティ定義を取得する。Graph Retrieval はリンクされたエンティティから k-hop の範囲内にあるトリプレットを取得する。HG Retrieval はリンクされたエンティティを含むハイパーエッジをパターンの優先度に基づいて取得する。

D 事例分析

表 6 に HG Text と Proposed の事例比較を示す。事例 1 はマルチモーダル情報が有効な場合, 事例 2 は検索失敗の場合である。事例 1 では, HG Text は作用機序を取得できるが, テキスト化された知識のみを使用するため L-DOPA と D-DOPA を区別できず失敗した。Proposed は SMILES 埋め込みによりキラリティ (@@) を直接利用し, 光学異性体を区別

して成功した。事例 2 では, 検索段階で入力中の dermatitis ではなく適応症 (alcohol dependence) が優先取得されたため, 両手法とも失敗した。これは性能が検索品質に依存することを示している。

表 4 ハイパーエッジパターン別統計

パターン	数	平均ノード	スコア
作用機序	14,764	5.00	1.0
パスウェイ	1,625	51.59	0.8
薬剤ハブ	2,105	21.43	0.6
疾患ハブ	1,917	76.12	0.6
曝露ハブ	2,083	8.64	0.6
PPI コミュニティ	786	33.57	0.5
化学構造類似性	1,282	2.98	0.5
解剖学的階層	14,032	2.00	0.4
生物学的プロセス	13,630	4.88	0.4
疾患階層	5,669	6.68	0.4
表現型階層	5,041	4.72	0.4
分子機能階層	2,021	7.72	0.4
細胞構成要素	972	5.98	0.4
エンティティ定義	121,673	1.23	0.0
合計	187,600	6.49	-

表 5 検索コンテキストの比較

入力	Famotidine-associated delirium. A series of six cases. Famotidine is a histamine H2-receptor antagonist...
Text	Famotidine: competitive H2 receptor antagonist. Cyclizine: H1 antagonist for vomiting. Cyproheptadine: serotonin antagonist.
Graph	Famotidine→HRH2, CYP1A2, SLC22A2, SLC22A6, SLC22A8, SLC47A1. Confusion→Delirium. (50 nodes, 48 edges)
HG	[drug_hub] Famotidine→CYP1A2, HRH2, SLC22A6... (6 proteins). [phenotype_hier] Confusion→Delirium. (16 nodes, 4 hyperedges)

表 6 事例分析: HG Text vs Proposed

事例 1: マルチモーダル情報が有効	
入力	Levodopa-induced dyskinesia among multiethnic Malaysians with Parkinson's disease.
取得知識	[triad] Levodopa→DDC→Parkinson's disease. [triad] Levodopa→DRD5→mental disorder. [entity_def] SMILES: N[C@@H](Cc1ccc(O)c(O)c1)C(O)=O
HG Text	テキストのみでは L/D-DOPA を区別不可。予測: ×
Proposed	SMILES 埋め込みでキラリティ (@@) を反映。予測: ○
事例 2: 検索失敗	
入力	Phenytoin-induced dermatitis in epileptic patients.
取得知識	[triad] Phenytoin→COMT→alcohol dependence. [triad] Phenytoin→CYP2D6→alcohol dependence. [triad] Phenytoin→CYP2E1→alcohol dependence.
HG Text	dermatitis ではなく適応症が取得された。予測: ×
Proposed	同一サブ HG のため検索失敗の影響。予測: ×