信頼できる知識グラフ推論のための評価リソースの自動構築

井田龍希　三輪誠
豊田工業大学大学院
{sd24501,makoto-miwa}@toyota-ti.ac.jp

概要

知識グラフ推論では，グラフ上の膨大な複数のパスを考慮した信頼できる推論が必要である．このような推論モデルの実現には，開発・評価するためのリソースが必要であり，LLM を活用した自動構築が有望である．しかし，LLM はパスに忠実でも，単なる相関関係などを含む論理的に弱い論証を生成し，根拠が結論を支持する論理的強度の度合い（以下，健全性）が高い推論との区別ができない．そこで，本研究では知識グラフ推論におけるパスに基づく論証生成タスクを定式化し，パスに接地した論証と健全性スコアで構成される開発・評価リソース ArgLink を自動構築する．検証では，健全性スコアが人間の健全性判断と高い相関を示した．

1　はじめに

知識グラフ（KG）は構造化知識の表現基盤として重要であり，その補完タスクである知識グラフ推論は実世界応用において重要である [1, 2]．KGE や GNN などの既存手法 [3, 4, 5] が高い性能を達成する一方で，近年では LLM を活用する研究 [6, 7] が注目を集めている．しかし，LLM を活用した既存手法の多くはクエリ近傍の事実のみを用いた局所的かつ断片的な推論にとどまり，より広い文脈から情報を統合できる LLM の能力を十分には活用できていない．信頼性の高い知識グラフ推論のためには，膨大な候補パスから一貫した論証を合成する能力が不可欠である（図 1 参照）．

しかし，このような高度な知識グラフ推論モデルを開発・評価するためのリソースは現在のところ存在しない．膨大な候補パスの分析に基づく高品質な論証の手動での作成はコストの観点から困難であり，その自動構築が期待される．そこで本研究では，既知の正しい事実 (h, r, t) の成立根拠となり得る候補パスを収集し，それらに基づき LLM に論証を生成させる論証生成タスクを定式化し，データ構築の自動化を実現する．本タスクは，既存の知識グラフに基づく質問応答（KGQA）[8, 9] のように回答への構造化されたパスの発見と検証を目的とするのではなく，KG 内の事実が欠落した状況において，広範な候補パスから成立根拠を分析し，説得力のある論証を構築することに主眼を置くものである．

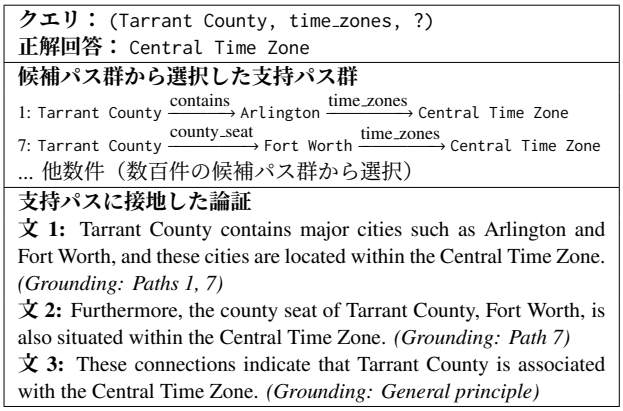

図 1　ArgLink の具体例．クエリとその正解に対応する膨大な候補パス群から，モデルは支持パス群を選択し，各文がパスに接地した一貫した論証を生成する．

この LLM を用いた自動構築では，論理的に強固な，すなわち健全性が高い論証と単なる相関に基づく弱い論証の区別が困難になる課題がある．KG には論理的に強固な証拠だけでなく，国籍推定における卒業大学の所在地といった希薄な相関関係しか存在しない場合も多い．LLM は，こうした弱い証拠に即して，忠実であるが論理的に弱い論証を生成する傾向があり，これが推論の確信度を不明瞭にする．信頼できる知識グラフ推論のためには，単なるグラフへの忠実性の担保だけでなく，論証の論理的強度を評価・区別できる新たな基準が不可欠である．

この課題に対応するため，本研究では，LLM に二つの独立した役割を与えて，生成した論証の健全性を自動評価するフレームワークを提案する．まず，論証生成 LLM が候補パスに基づき論証を生成し，続いて評価 LLM がその論理的強度を健全性スコアとして評価する．このフレームワークに基づき，各

文が根拠となるパスに接地した論証とその健全性スコアを併せ持つ大規模リソース ArgLink を自動構築する．ArgLink は，知識グラフ推論における推論過程とその健全性の評価への利用が期待できる．

本研究の貢献は以下の 3 点である．

1. **リソース作成における課題の特定**：LLM は KG 内の事実には忠実であっても論理的な強さを考慮しないため，健全性が低い論証を生成する傾向があることを特定した．
2. **健全性評価を含む自動論証生成手法**：論理的健全性を自動的に定量化する解決策として，論証生成 LLM と健全性評価 LLM からなる自動評価フレームワークを用いた健全性評価を含む自動論証生成手法を提案した．
3. **検証済みの評価リソースの提供**：推論の論理的強度を区別可能なモデル開発を支援する，人間の健全性スコアの判断と強い相関（スピアマンの相関係数 $\rho = 0.786$）を持つ初の大規模リソース ArgLink を自動構築した．

2 関連研究

本研究は知識グラフ推論およびその説明可能性を対象とし，人間が読める論証を生成する．その動機は既存の事後的な説明可能な AI（XAI）とは異なり，推論プロセスの言語化自体が予測の精度や信頼性の向上に寄与するという仮説に基づいている．この立ち位置を踏まえ，本節ではまず知識グラフ推論モデルの変遷（2.1 節）と知識グラフ推論の説明可能性に関する既存手法（2.2 節）について概観する．

2.1 知識グラフ推論モデルの変遷

KGE [1, 10, 11] や GNN [12] は，知識グラフ推論において構造表現の学習に焦点を当ててきた．最近では，事前学習済み言語モデルを知識グラフ推論に活用する手法がある．KG-BERT [13] は知識グラフ推論をテキスト分類問題として再構成し，SimKGC [14] は効率的な対照学習により性能を向上させた．さらに大規模な LLM の登場により，KG タスクのテキスト化が進んでいる．LLM を直接的なエンドツーエンドの推論器として用いる方針では，推論パスに基づくファインチューニング [6] や，GS-KGC [7] のように局所的な部分グラフ情報でプロンプトを強化する手法が検討されている．また，エンティティや関係を特殊トークンとして LLM を活用する手法 [15] や既存 KGE モデルの性能向上のために LLM を活用する手法 [16] も提案されている．

2.2 知識グラフ推論における説明可能性

知識グラフ推論の説明可能性に関する研究は，記号的な出力から言語的な論証へと進展した．まず，推論のプロセスがモデルの動作から直接読み取れる手法として，グラフから直接人間が読めるホーン節を発見する AMIE+ [17] や，強化学習を用いて重要なパスを探索する MINERVA [18] などが挙げられる．一方で，事後的な説明のアプローチとして，GNNExplainer [19] のように特定の予測に対して，大きな影響を与えるコンパクトな部分グラフを特定する手法も存在する．さらに最近の研究では，LLM を用いた知識グラフ推論の自然言語での説明生成が進んでいる．近年，生成された説明の品質を評価するために “LLM as a Judge” [20] が広く採用されている．

3 提案手法

本節では，提案する論証生成タスクの定式化および，その評価リソース ArgLink の自動構築手法を詳述する．パイプラインの概要を図 2 に示す．まず 3.1 節では，健全性スコアを含む新たなタスク定義を行う．続いて 3.2 節では，証拠リークを防ぎつつ候補パスを抽出する入力データ構築を説明する．最後に 3.3 節では，LLM による論証生成とその健全性を自動評価するフレームワークについて述べる．

3.1 論証生成タスクの定式化

本研究では，論証生成タスクを新たに定式化する．クエリ $(h, r, ?)$ に対して，正解となるターゲット回答の集合を $T_{target} = \{t_1, \ldots, t_k\}$，候補パス集合を $P = \{p_1, \ldots, p_n\}$ とする．本タスクの目的は，T_{target} に含まれる全回答の成立根拠を包括的に説明する一貫した論証 A とその全体的な論理的強度を表す健全性スコア $s \in \{1, \ldots, 5\}$ のペア (A, s) を生成することである．ここで，論証 A は文単位での証拠への接地と自由な記述を両立するため，m 個の構成要素[1)]からなる系列 $A = [c_1, \ldots, c_m]$ として定義する．各構成要素 c_i は，タプル (a_i, T_i, E_i) であり，a_i は論証を構成する i 番目の文，$T_i \subseteq T_{target}$ はその文が説明対象とする（一つまたは複数の）回答の集合，E_i は文の証拠となる支持パス集合 $E_i \subseteq P$ または一般原則である．最終的な論証は，系列内の文

1) m は論証により異なる．

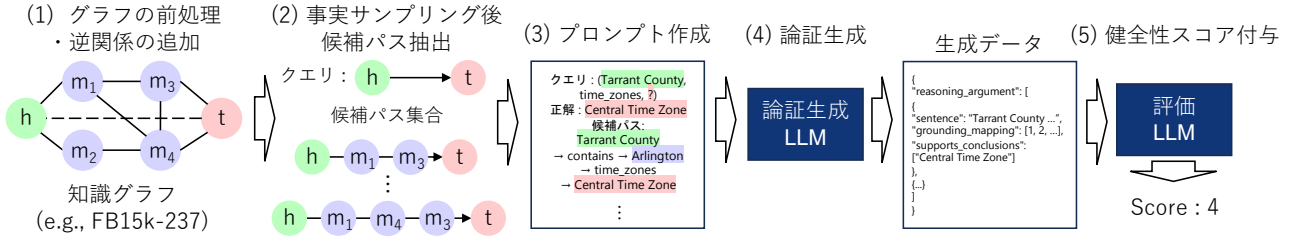

図 2　ArgLink 生成パイプラインの概要．KG を用いて以下の処理を行う．(1) 双方向のパス探索を可能にするため，逆関係を追加する．(2) 事実をサンプリングし，深さ優先探索を用いて候補パスを抽出する．この際，証拠リークの防止のため，ターゲット回答への直接的な辺を削除する．(3) クエリと候補パスを含むプロンプトを構築する．(4) 論証生成 LLM を用いて，検証可能な構造化された論証を生成する．(5) 評価 LLM を用いて論証に健全性スコアを付与する．

$a_1, \ldots, a_m$ をこの順序で連結して形成される．この定式化により，クエリの複数の回答への推論過程を構造的に保持しつつ，全体として統合された論証の構築が可能となる．また，健全性スコア s により，論理的に強固な論証と単なる相関に基づく弱い論証を区別でき，推論の信頼性が向上する．

3.2　入力データ構築

ArgLink 自動構築の第一段階として，入力事例を準備する．各事例は，推論対象となる事実とその成立根拠として利用され得る候補パス群から成る．

クエリのサンプリングとパス抽出　FB15k-237 [21] と WN18RR [22] を基盤とし，ArgLink-FB15k-237 と ArgLink-WN18RR を構築する．まず，それぞれから 5,000 のクエリをサンプリングした．各クエリの正解回答集合 T をターゲット回答 T_{target} と推論の入力として利用する既知回答 T_{known} に分割した上で，広範なパス発見のため既存のすべての辺に対して逆関係を追加する．そして，深さ優先探索を用いて，ヘッド h と各ターゲット回答 $t \in T_{target}$ 間の単純パスを探索する．

証拠リークの防止　候補パスからの論証生成のためには，推論対象の事実が候補パス内に含まれる情報リークを防ぐ必要がある．パス抽出では，ターゲット回答 $t \in T_{target}$ への直接的な証拠（順方向辺 (h, r, t) および逆方向辺 (t, r_{inv}, h)）をグラフから削除する．これにより，単なる回答の存在確認ではなく，周囲の候補パスからの推論（例えば，クエリ（**オバマ**, **配偶者**, ?）の推論で，**オバマ** $\xrightarrow{\text{子供}}$ X $\xrightarrow{\text{母親}}$ Y といったパスを参照）が必要となる．

3.3　自動評価する論証生成手法

本手法では，論証生成 LLM と評価 LLM から成る二段階の LLM パイプラインを用いる．

3.3.1　論証生成 LLM による論証生成

本ステップではまず，クエリ，ターゲット回答 T_{target}，候補パス群 P に加え，ヘッド h の説明文や関係のセマンティクスを理解するための事例を含むプロンプトを構築する．論証生成 LLM の役割は，候補パス群 P から各回答 $t \in T_{target}$ の証拠となる支持パスを選別し，それらに基づく説得力のある一貫した論証 A を構築することである．パスが論証として弱いあるいは単に相関的である場合でも，モデルは慎重かつ正確な表現（例：「〜によって可能性が示唆される」）を用いて，与えられた証拠の範囲内で最大限の論理的な説明を行うように指示される．このプロセスにおいて，論証生成 LLM は論証を個別の文 a_i の系列として生成すると同時に，各文に対応する説明対象の回答 $t_i \in T_{target}$ およびその支持パス E_i を紐付けた構造化された出力を行う．ここで，モデルは単に各文を羅列するのではなく，一貫した論理的な論証を生成することが求められる．この形式により，推論過程を各証拠へ直接的に接地させることが可能となり，生成された論証の忠実性を客観的に検証可能な状態にする．

3.3.2　評価 LLM による健全性評価

論証生成 LLM が論証を生成した後，別の評価 LLM が包括的な 5 段階評価ルーブリックに基づき，忠実性を前提とした論理的健全性を評価する．具体的には，まず主張が提供されたパスにのみ基づいているか（忠実性）を検証し，その上で，結論を支持する証拠の論理的強度を評価軸としてスコアを決定する．評価 LLM のタスクは，判断の理由を含むフィールドと，1（Unusable）から 5（Excellent）の 5 段階からなる全体的な健全性スコアを含む JSON を出力することである．このスコアは，論理的に強固で健全な論証と，事実に忠実でありながら論理的に弱い論証を区別するためのシグナルとなる．

表 1 評価 LLM スコアと人手評価スコアの相関 (n=50)

評価 LLM モデル	Spearman's ρ	Kendall's τ
Gemini 2.5 Flash Lite	0.786	0.678
Claude Haiku 4.5	0.786	0.704
GPT-4o	0.763	0.676

4 人手評価による検証

本節では，評価 LLM が付与した健全性スコアの信頼性を検証するため，人手評価との相関分析および事例研究を行う．

4.1 人手評価との相関分析

自動生成された健全性スコアの信頼性を検証するため，FB15k-237 を基盤として作成した ArgLink-FB15k-237 から 50 件の事例をランダムにサンプリングし，人手による評価を行った．2 名のアノテータが評価 LLM と同じ 5 段階基準で健全性スコアを評価したところ，アノテータ間の一致率は Krippendorff's $\alpha = 0.877$ に達し，評価ルーブリックの明確性と信頼性を確認した．

次に，人間評価の中央値と評価 LLM のスコアの相関を表 1 に示す．データ構築に用いた Gemini 2.5 Flash Lite は人間の判断と強い正の相関（スピアマンの相関係数 $\rho = 0.786$）を示した．さらに，異なるモデル（Claude Haiku 4.5, GPT-4o）を評価 LLM として用いた場合でも，同様に高い相関が見られた．この結果は，LLM が自身や同系列モデルの出力を過大評価する傾向にある自己選好バイアスの影響が限定的であり，ArgLink に付与された健全性スコアの論理強度の指標としての有効性を示唆している．

4.2 事例研究による定性分析

評価 LLM が付与した健全性スコアの特徴を具体的に確認するため，事例研究による定性的な分析を行った（詳細は付録 A 参照）．例えば，地理的な包含関係に基づく論理的な推論には高いスコアを付与する一方，共通の受賞歴といった希薄な相関関係に依存する論証には，低いスコアを付与していた．これは，ArgLink が論理的健全性の評価リソースとして有効に機能していることを裏付けている．

5 ArgLink の特性分析

本節では，自動構築した ArgLink の特性分析を行う．詳細な分析結果は付録 B と C にて詳述する．

表 2 ArgLink の統計情報

項目	ArgLink-FB15k-237	ArgLink-WN18RR
総事例数	4,876	4,969
平均文数	4.58	4.29
平均候補パス	415.76	102.47
平均支持パス	10.25	5.96

表 3 ArgLink における評価 LLM の健全性スコア分布

健全性スコア	ArgLink-FB15k-237	ArgLink-WN18RR
5 (Excellent)	35.02%	38.06%
4 (Good)	16.74%	12.12%
3 (Acceptable)	16.82%	13.38%
2 (Poor)	18.97%	19.58%
1 (Unusable)	12.45%	16.86%

5.1 統計的特性

FB15k-237 と WN18RR を基盤として作成した ArgLink-FB15k-237 と ArgLink-WN18RR の統計を表 2 に示す．候補パス群と最終的に選択された支持パス群の数には大きな差があり，このことから LLM は膨大なパス群から推論のために重要なパスのみを選別していることが確認できる．

5.2 健全性スコアの分析

健全性スコアの分布を表 3 に示す．表より知識グラフのクエリの約 50%が，間接的・相関的・論理的に弱いパスを用いて説明せざるを得ず，低い健全性スコアにつながることが明らかになった．これは，すべての正しい事実が強力または直接的な証拠を持つわけではないという知識グラフ推論の現実を反映している．ArgLink の価値は，この証拠の不確実性を定量化している点にある．

6 おわりに

知識グラフ推論において，根拠となるパスに接地した論証生成を可能にするリソースの自動構築手法を提案し，初の大規模リソース ArgLink を構築した．論証生成 LLM と評価 LLM を用いた自動評価フレームワークにより，論理的健全性の評価を含むデータセットの自動構築を実現した．分析の結果，標準的なデータセットにおける事実の約半数が，論理的に強力な証拠ではなく，相関関係などの比較的弱い証拠に依存していることが明らかになった．また，自動に付与された健全性スコアが人間の判断と強く相関することを示し，信頼できる推論モデルの開発における有効な指標として利用できることを示した．

参考文献

[1] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In **Advances in Neural Information Processing Systems 26**, pp. 2787–2795, 2013.

[2] Shivani Choudhary, Tarun Luthra, Ashima Mittal, and Rajat Singh. A survey of knowledge graph embedding and their applications. **arXiv preprint arXiv:2107.07842**, 2021.

[3] Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal A. C. Xhonneux, and Jian Tang. Neural Bellman-Ford Networks: a general graph neural network framework for link prediction. In **Advances in Neural Information Processing Systems 34**, pp. 29476–29490, 2021.

[4] Zhaocheng Zhu, Xinyu Yuan, Michael Galkin, Louis-Pascal A. C. Xhonneux, Ming Zhang, Maxime Gazeau, and Jian Tang. A*Net: a scalable path-based reasoning approach for knowledge graphs. In **Advances in Neural Information Processing Systems 36**, 2023.

[5] Zhanke Zhou, Yongqi Zhang, Jiangchao Yao, Quanming Yao, and Bo Han. Less is More: one-shot subgraph reasoning on large-scale knowledge graphs. In **Proceedings of the International Conference on Learning Representations**, 2024.

[6] Dong Shu, Tianle Chen, Mingyu Jin, Chong Zhang, Mengnan Du, and Yongfeng Zhang. Knowledge graph large language model (KG-LLM) for link prediction. In **Proceedings of the Asian Conference on Machine Learning**, pp. 143–158, 2024.

[7] Rui Yang, Jiahao Zhu, Jianping Man, Hongze Liu, Li Fang, and Yi Zhou. GS-KGC: A generative subgraph-based framework for knowledge graph completion with large language models. **Inf. Fusion**, Vol. 117, p. 102868, 2025.

[8] Linhao Luo, Yuan-Fang Li, Gholamreza Haffari, and Shirui Pan. Reasoning on Graphs: faithful and interpretable large language model reasoning. In **Proceedings of the International Conference on Learning Representations**, 2024.

[9] Linhao Luo, Zicheng Zhao, Gholamreza Haffari, Yuan-Fang Li, Chen Gong, and Shirui Pan. Graph-constrained Reasoning: Faithful reasoning on knowledge graphs with large language models. In **Proceedings of the International Conference on Machine Learning**, pp. 41540–41565, 2025.

[10] Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. In **Proceedings of the International Conference on Learning Representations**, 2015.

[11] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In **Proceedings of the International Conference on Machine Learning**, pp. 2071–2080, 2016.

[12] Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In **Proceedings of the Extended Semantic Web Conference**, pp. 593–607, 2018.

[13] Liang Yao, Chengsheng Mao, and Yuan Luo. KG-BERT: BERT for knowledge graph completion. **arXiv preprint arXiv:1909.03193**, 2019.

[14] Liang Wang, Wei Zhao, Zhuoyu Wei, and Jingming Liu. SimKGC: Simple contrastive knowledge graph completion with pre-trained language models. In **Proceedings of the Annual Meeting of the Association for Computational Linguistics**, pp. 4281–4294, 2022.

[15] Lingbing Guo, Zhongpu Bo, Zhuo Chen, Yichi Zhang, Jiaoyan Chen, Yarong Lan, Mengshu Sun, Zhiqiang Zhang, Yangyifei Luo, Qian Li, Qiang Zhang, Wen Zhang, and Huajun Chen. MKGL: mastery of a three-word language. In **Advances in Neural Information Processing Systems 38**, 2024.

[16] Pengcheng Jiang, Lang Cao, Cao Xiao, Parminder Bhatia, Jimeng Sun, and Jiawei Han. KG-FIT: knowledge graph fine-tuning upon open-world knowledge. In **Advances in Neural Information Processing Systems 38**, 2024.

[17] Luis Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. Fast rule mining in ontological knowledge bases with AMIE+. **VLDB J.**, Vol. 24, No. 6, pp. 707–730, 2015.

[18] Rajarshi Das, Shehzaad Dhuliawala, Manzil Zaheer, Luke Vilnis, Ishan Durugkar, Akshay Krishnamurthy, Alex Smola, and Andrew McCallum. Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning. In **Proceedings of the International Conference on Learning Representations**, 2018.

[19] Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. GNNExplainer: generating explanations for graph neural networks. In **Advances in Neural Information Processing Systems 32**, pp. 9240–9251, 2019.

[20] Haitao Li, Qian Dong, Junjie Chen, Huixue Su, Yujia Zhou, Qingyao Ai, Ziyi Ye, and Yiqun Liu. LLMs-as-Judges: A comprehensive survey on LLM-based evaluation methods. **arXiv preprint arXiv:2412.05579**, 2024.

[21] Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon. Representing text for joint embedding of text and knowledge bases. In **Proceedings of the Conference on Empirical Methods in Natural Language Processing**, pp. 1499–1509, 2015.

[22] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2D knowledge graph embeddings. In **Proceedings of the AAAI Conference on Artificial Intelligence**, pp. 1811–1818, 2018.

A　事例研究の詳細

評価 LLM がどのように健全な論証と弱い論証を区別しているかを確認するため，健全性スコアが高い事例と低い事例を定性的に比較する．

高スコア事例：　図 1 に示した事例では，Tarrant County のタイムゾーンを推論するために，「Arlington や Fort Worth を地理的に含んでいる（contains）」という事実と，それらの都市が「Central Time Zone にある」という事実を組み合わせている．これは部分が全体と同じ属性を持つという包含関係に基づいており，論理的に強固である．評価 LLM はこの推論の強さを評価し，最高のスコア 5 を付与した．

低スコア事例：　一方，図 3 に示す映画監督 Nora Ephron の出身校 Beverly Hills High School を推論する事例では，利用可能なパスが共通の受賞歴といった間接的な情報に限定された．生成された論証は，「Nora Ephron は Blake Edwards と関連があり，彼は Beverly Hills High School の卒業生である」という内容であったが，生成された論証の中でもこれらのつながりは間接的であり，決定的ではないと明記されている．これは，パス自体には忠実であっても，論理的には単なる相関関係に過ぎない．評価 LLM はこうした論理の弱さを正しく評価し，スコア 2 を付与した．この事例は，ArgLink が論証の事実的整合性だけでなく，その論理的強度を区別できるリソースであることを示している．

クエリ： (Nora Ephron, educational_institution, ?)
正解回答： Beverly Hills High School

候補パス群から選択した支持パス群

1: Nora Ephron $\xrightarrow{\text{nominated}}$ Razzie Award → ...
→ Blake Edwards $\xrightarrow{\text{alumnus}}$ Beverly Hills High School
7: Nora Ephron → ... Jewish people → ...
→ Tori Spelling → ... → Beverly Hills High School
... 他数件（数百件の候補パスから選択）

支持パスに接地した論証

文 1: Nora Ephron is linked to Blake Edwards, who is documented as a student and graduate of Beverly Hills High School. *(Grounding: Path 1)*

文 2: Furthermore, she shares connections with individuals like Tori Spelling, who are also graduates of the school. *(Grounding: Paths 7, ...)*

文 3: These indirect associations suggest a potential, though **not definitive**, link between Nora Ephron and Beverly Hills High School. *(Grounding: General principle)*

図 3　健全性が低い論証の事例．モデルは利用可能なパスから論証を構築するが，その推論が相関的であり決定的ではない（not definitive）ことを正しく自動評価している．

B　構造パターンの詳細分析

論証 LLM が選択した支持パスの構造パターンを分析するため，論証内で引用された支持パスに対応するメタパス（関係の連なり）を集約した．代表的な関係タイプにおける主要な構造パターンを表 4 に示す．ある関係タイプを持つ全事例のうち，特定のメタパスが証拠として採用された割合を網羅率として定義した．これらの網羅率が高い構造パターンは，モデルがある関係タイプに対して重要な証拠を一貫して特定・利用していることを示唆している．例えば，関係/music/genre/artists においては，72.7%の事例で親ジャンルを経由するメタパスが証拠として採用されている．例えばバロック・ポップのアーティストを推論する際に，親ジャンルであるポップ・ミュージックを経由するパスを参照しており，これは人間の直感とも合致する論理的に健全な推論パターンであるといえる．

表 4　ArgLink-FB15k-237 における代表的な関係タイプの主要な構造パターン．

関係タイプ	構造パターン	網羅率
/music/genre/artists	/music/genre/parent_genre → /music/genre/artists	72.7%
/location/location/ time_zones	/location/location/contains_inv → /location/location/time_zones	70.0%
/music/record_label/ artist_inv	/music/genre/artists_inv → /music/genre/artists → /music/record_label/artist_inv	65.2%

C　関係タイプによる健全性スコア

健全性スコアの決定要因を調査するため，関係タイプ別の平均スコアを算出した結果，以下の特徴が明らかになった．

健全性スコアが高い関係タイプ：　米国の地域区分とそれが属する郡の関係などが該当する．これらは，地理的な包含関係を補完する代替パス（例：contains など）がグラフ内に存在し，演繹的な裏付けが容易なため，一貫して高いスコアを記録した．

健全性スコアが低い関係タイプ：　スポーツリーグとその所属チームの関係などが該当する．これらは，事実は存在するものの，周囲に直接的な因果を示す証拠が少なく，相関的なパスに依存した論理的な飛躍を含む論証になりやすいため，低いスコアにとどまった．