

一般発表 | Track 3: 情報検索・情報推薦・ソーシャルメディア

2026年3月1日(日) 13:00 ~ 15:10 | 会場

[5G] 推薦システム(グラフ/知識)

座長:林 康弘(武蔵野大学) コメントータ:有次 正義(熊本大学) Jコメントータ:吉丸 直希(同志社大学) ジュニアコメントータ:森部 想水(九州大学)

13:00 ~ 13:25

[5G-01] マルチドメイン推薦のための協調知識グラフとGNNによる知識転移手法

*山路 大樹¹ (1. 株式会社NTTドコモ)

13:25 ~ 13:50

[5G-02] グラフ構造を考慮した負例推定に基づく推薦モデルの提案

*曾根 良行¹、道瀬 悠磨¹、原 隆浩¹ (1. 大阪大学)

13:50 ~ 14:15

[5G-03] グループ単位の関連度評価に基づく知識グラフ探索アルゴリズムの設計と評価

*八尾 正剛¹、福井 健太¹、Li Guangcan¹、鬼塚 真¹ (1. 大阪大学)

14:15 ~ 14:40

[5G-04] GLCSRec: Integrating Graph Collaborative Signals with Large Language Models for Item Cold-Start Recommendation

*Lin Ying¹、Chen Chongxian¹、Fan Xin¹、Hayato Yamana¹ (1. 早稲田大学)

14:40 ~ 15:05

[5G-05] 適応的回転ゲートと局所探索を用いた量子ビット進化アルゴリズムによる TGNN の説明手法

*小原 陸哉¹、佐々木 勇和¹、小島 豊¹ (1. 大阪大学)

マルチドメイン推薦のための協調知識グラフと GNNによる知識転移手法

山路 大樹[†]

[†] 株式会社 NTT ドコモ 〒100-6150 東京都千代田区永田町 2-11-1

E-mail: [†]daiki.yamjai.rk@nttdocomo.com

あらまし マルチドメイン推薦は、複数ドメインにまたがるユーザ行動データを活用することで、ユーザ嗜好をより深く理解すると同時に、データが乏しいドメインにおける推薦精度を向上させるための重要な推薦技術である。しかし実環境においては、複数ドメインにわたってアイテムを利用するユーザは限定的であることが多く、ドメイン間の知識転移による恩恵を十分に享受できないという構造的課題が存在する。この課題を解決するために、複数ドメインに属するユーザ・アイテム間のインタラクションから構成されるグラフに、ドメイン横断的なアイテム間の意味的關係を表現する知識データ（知識グラフ）を統合した協調知識グラフを構築する。そして、各ドメイン固有の情報を保持しつつ、ドメイン不変な知識を効果的に転移する新たな知識転移手法である DKGAT（Disentangled Knowledge Graph Attention Network）を提案する。Amazon Review Data を用いた実験により、提案手法が既存手法を上回る推薦精度を達成すること、アブレーションスタディおよび可視化分析を通じて、提案手法が有効であることを示す。

キーワード グラフデータ、知識グラフ、グラフニューラルネットワーク、推薦モデル、リンク予測

1 序 論

マルチドメイン推薦（Multi-domain Recommendation: MDR）は、書籍、映画、音楽、家電など、複数の異なるドメインにまたがるアイテムを対象として、ユーザの嗜好を統合的に捉える推薦技術である。代表例として、EC サイトにおけるクロスセル（買い合わせ）推薦や、ニュース閲覧履歴を起点とした動画コンテンツの推薦などが挙げられる。

MDR の重要な利点は、複数ドメインの行動データを統合的に分析することで、単一ドメインでは観測しにくいユーザの潜在的嗜好をより精緻に捉えられる点にある [1] [7] [27]。特に、あるドメインで十分な行動履歴が得られない場合でも、他ドメインの情報を補助的に利用することで、推薦精度の向上が期待できる [8] [15]。この背景には、「ユーザの根源的な嗜好はドメインを超えて部分的に共有される」という仮説が存在し、データが豊富なドメインから得られた知識を他のドメインへ転移する知識転移（knowledge transfer）を通じて、推薦品質の向上が期待される。

MDR において、ドメイン間で共有される嗜好を捉える上で重要な役割を担うのが、複数ドメインにまたがって行動履歴を持つ横断ユーザ（overlapping users）である。横断ユーザは、ドメイン間の嗜好の対応関係を結び付ける橋渡しとして機能する。例えば、映画ドメインで SF ジャンルを好むユーザが、書籍ドメインでも硬派な SF 小説を好む場合、横断ユーザを介してこのような嗜好の相関関係を学習可能である。しかし実環境では、横断ユーザは少数に限られており、横断ユーザのみに依存した知識転移では、MDR の効果が制限されるという課題がある。

この課題に対して、ユーザ・アイテムのインタラクションデータに加え、アイテム属性や外部知識を活用する推薦手法が研究さ

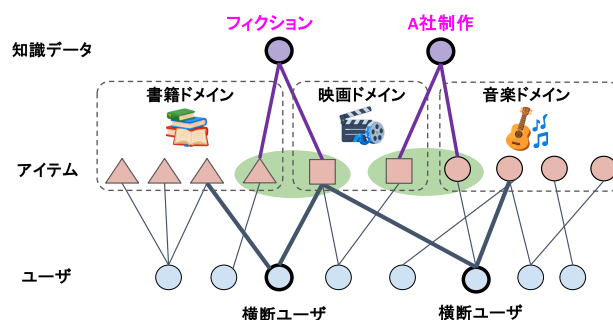


図 1: 複数ドメインを含む協調知識グラフの例。知識データを導入することで、横断ユーザのみに依存せず、共有知識を介したドメイン横断的な嗜好推定が可能となる。一方で、ドメイン不変な知識とドメイン固有の知識を区別せずに共有すると、負の転移や知識グラフ由来のノイズが他ドメインへ伝播するおそれがある。

れてきた [14] [4]。これらの手法は、アイテムのメタデータを潜在表現へ取り込むことで、ドメイン横断的な行動情報が乏しい場合でもメタデータを活用した推薦精度の向上が期待できる。さらに近年では、アイテム属性や背景知識を知識グラフ（Knowledge Graph: KG）として構造化し、ユーザ・アイテムのインタラクションからなるグラフと統合した協調知識グラフ（Collaborative Knowledge Graph: CKG）を構築したうえで、グラフニューラルネットワーク（Graph Neural Network: GNN）を適用する研究が進んでいる [21] [24] [29] [25] [30] [28] [33]。特にドメインが複数存在する場合に、ドメインをまたがって共有される属性や概念を知識グラフとして表現することにより、横断ユーザが少ない場合でも、知識エンティティを介した GNN に

よるドメイン間の情報伝播が可能になる。例えば、書籍と映画が共通して「フィクション」や「ミステリー」といったジャンル概念に接続されていれば、ドメイン横断的な行動情報が乏しくても、こうした共有知識を通じてドメイン横断的な嗜好の対応を学習できる可能性がある（図1）。

一方で、MDR において協調知識グラフを扱う場合、以下の二つの課題がある。

1. **負の転移 (Negative Transfer)** : ユーザの意思決定には、複数ドメインで共有されるドメイン不変の嗜好（例：映画も書籍もフィクション好き）と、各ドメインに依存するドメイン固有の嗜好（例：映画ではホラーを好むが、書籍ではエンタメ性を重視する）が混在している。これらを区別せずに共有表現として学習すると、あるドメインに特有な嗜好やバイアスが他ドメインへ不適切に転移し、推薦精度が低下する可能性がある。
2. **知識グラフのノイズや有用性の定義の違い**: 実世界の知識グラフには、推薦に有用でない関係や、粗い属性、不完全・曖昧な関係が含まれることがある。これらを無差別に利用すると、アイテム表現にノイズが混入し、推薦性能の低下を招きうる。また、ある知識がドメイン不変の嗜好推定には有用でも、ドメイン固有の嗜好推定には有用でない場合もあり、利用すべき知識は一様ではない。

以上を踏まえ、本研究では、マルチドメインの協調知識グラフに対して、ドメイン間で共有すべき知識と各ドメインに保持すべき知識を分離しながら学習する GNN ベースの推薦手法 DKGAT (**D**isentangled **K**nowledge **G**raph **A**ttention Network) を提案する。DKGAT では、ユーザ・アイテム・知識エンティティに関する表現を、ドメイン不変表現とドメイン固有表現に明示的に分離する。不変表現は複数ドメインで共有される共通嗜好の表現を担い、固有表現は各ドメインの文脈に依存する嗜好を保持する。さらに、知識グラフ上にアテンション機構と重み付き知識グラフ埋め込みを導入することで、推薦に有用な関係を選択的に強調し、ノイズの影響を抑制する。

本研究の主な貢献は以下の三点である。

- **マルチドメイン協調知識グラフに対する不変/固有分離学習の提案**: ユーザ・アイテムの行動表現だけでなく、知識グラフ由来の表現もドメイン不変表現とドメイン固有表現に分離して学習する枠組みを提案する。これにより、横断ユーザが少ない状況でも、共有すべき知識と各ドメイン固有の知識を区別した知識転移を可能にする。
- **知識選択機構を備えた二系統 GNN フレームワークの設計**: ドメイン不変表現とドメイン固有表現それぞれに対して独立した GNN を適用する二系統のフレームワークを設計した。これにより、ドメイン間で共有される嗜好とドメイン固有の嗜好を同時にモデル化した推薦を可能にする。さらに、知識グラフ上においても二系統のアテンション機構と重み付き知識グラフ埋め込みを導入することで、推薦に有用な知識関係を不変表現および固有表現それぞれで選択的に強調し、かつノイズや不適切な知識共有の影響を抑制する。

- **公開データセットを用いた包括的検証**: Amazon Review Data を用いた複数ドメイン設定の実験により、提案手法が既存の MDR 手法および KG ベース推薦手法に対して高い性能を示すことを確認する。さらに、アブレーション、可視化、パラメータ分析を通じて、各構成要素の有効性を検証する。

2 関連研究

本研究に関連する研究として、(1) マルチドメイン推薦手法、(2) メタデータを活用した推薦手法、(3) 知識グラフと GNN を用いた推薦手法の三つに分類して整理する。

2.1 MDR 手法

マルチドメイン推薦 (MDR) は、複数ドメインのユーザ行動データを統合することで、ユーザ嗜好の推定の精緻化やデータスパース性の緩和を目的とした研究分野である。代表的な手法として、STAR [16], TreeMS [12], MGFN [31] などが提案されている。これらの手法では、ドメイン間で共有されるパラメータとドメイン固有のパラメータを明示的に設計することで、ドメイン間の知識共有を実現している。しかし、共有パラメータを通じて複数ドメインの勾配が同時に更新されるため、異なるドメインの目的が競合する「負の転移」が生じる可能性が指摘されている。この問題に対して、EDDA [11] は埋め込み表現を共有成分とドメイン固有成分に分離することで、ドメイン間の干渉を抑制する手法を提案している。また、負の転移を抑制するための手法 [9] [22] も提案されている。一方で、これらの MDR 手法の多くはユーザ・アイテムのインタラクションデータのみに基づいて学習を行うため、アイテムの背景知識や属性関係を活用したドメイン横断的な知識転移の可能性については十分に検討されていない。

2.2 メタデータを活用した推薦手法

ユーザ行動データのスパース性を補うために、アイテムの属性情報などのメタデータを活用する推薦手法も広く研究されている。Factorization Machines (FM) [14] や Neural FM (NFM) [4] は、ユーザ・アイテム・属性などの特徴量間の相互作用を潜在空間でモデル化する代表的な手法である。これらの手法では、共通の属性を通じて異なるアイテム間の関連性を捉えることができるため、ドメイン横断的な嗜好の推定や、ユーザ行動が少ない場合における嗜好の推定が可能である。しかし、これらの手法は主に特徴量間の組み合わせを扱うモデルであり、GNN のように高次の関係性や知識構造を十分に活用することができない。また、MDR 環境においてはドメイン間で共有される属性とドメイン固有の属性が混在するため、これらを区別せずに学習すると、ドメイン固有のバイアスが混入する可能性がある。

2.3 知識グラフと GNN を用いた構造的推薦

メタデータ間の複雑な関係を構造的に扱うため、アイテムの背景知識を知識グラフ (KG) として構築し、グラフニューラル

ネットワーク (GNN) を適用する研究が近年活発に行われている。KGAT [24] は、知識グラフ上の関係に対してアテンション機構を導入し、推薦に有用な関係を適応的に重み付けしながら高次の近傍情報を集約する手法である。さらに、KGCL [29] は対照学習を導入することで、知識グラフのノイズや不完全性に対する頑健性を向上させている。しかし、これらの手法の多くは単一ドメインの推薦シナリオを前提として設計されており、負の転移を考慮してドメイン間で共有すべき知識 (ドメイン不変知識) とドメイン固有の知識を区別して扱う枠組みは十分に検討されていない。そのため、MDR 環境において知識グラフを直接統合すると、ドメイン固有の情報が他ドメインへ不適切に伝播し、負の転移を引き起こす可能性がある。

3 問題定式化

3.1 インタラクシヨングラフ (Interaction Graph)

推薦シナリオにおけるユーザとアイテムのインタラクシヨンを、インタラクシヨングラフ $\mathcal{G}_{\text{int}} = (\mathcal{U}, \mathcal{I}, \mathcal{Y})$ として定義する。ここで \mathcal{U} は全ドメインにおけるユーザノードの集合、 \mathcal{I} は全ドメインにおけるアイテムノードの集合を表す。各エッジ $y_{ui} \in \mathcal{Y}$ は、ユーザ $u \in \mathcal{U}$ とアイテム $i \in \mathcal{I}$ の間で観測されたインタラクシヨンを表す。つまり、エッジの存在はユーザが当該アイテムに対して何らかの行動 (閲覧・購入・評価など) を行ったことを意味する。

3.2 知識グラフ (Knowledge Graph)

アイテムの属性情報や背景知識を保持する構造化データを、知識グラフ $\mathcal{G}_{\text{kg}} = (\mathcal{E}, \mathcal{R})$ として定義する。ここで \mathcal{E} はエンティティの集合であり、アイテムと知識データを含む ($\mathcal{I} \subset \mathcal{E}$)。 \mathcal{R} はエンティティ間の意味の関係 (relation) の集合である。知識グラフは、事実を表すトリプレットの集合 $\{(h, r, t) \mid h, t \in \mathcal{E}, r \in \mathcal{R}\}$ として記述される。例えば、(h : 『容疑者 X の献身』, r : ジャンル, t : ミステリー) といった構造が保持される。

3.3 協調知識グラフ (Collaborative Knowledge Graph)

インタラクシヨングラフと知識グラフを統合したグラフを、協調知識グラフ $\mathcal{G}_{\text{ckg}} = \{(h, r, t) \mid h, t \in \mathcal{E}', r \in \mathcal{R}'\}$ として定義する。ここで $\mathcal{E}' = \mathcal{U} \cup \mathcal{E}$ はユーザ・アイテム・知識エンティティを含むノード集合であり、 $\mathcal{R}' = \mathcal{Y} \cup \mathcal{R}$ はユーザとアイテムのインタラクシヨ関係および知識グラフの関係を含む関係集合である。このように構築された協調知識グラフでは、ユーザはアイテムノードを介して知識エンティティと間接的に接続されるため、行動データが観測されていないドメインに対しても知識データを通じた情報伝播が可能となる。特に、ジャンルやブランドといったドメイン横断的なエンティティを共有することで、横断ユーザが少ない状況下においても、ドメイン間の知識転移を実現できる。

3.4 マルチドメイン設定

マルチドメイン推薦の対象となるドメイン集合を $\mathcal{D} =$

$\{d_1, d_2, \dots, d_W\}$ とする。ここで W はドメイン数を表す。特定のドメイン $d_w \in \mathcal{D}$ におけるインタラクシヨンデータは、ユーザ集合 $\mathcal{U}^{d_w} \subseteq \mathcal{U}$ およびアイテム集合 $\mathcal{I}^{d_w} \subseteq \mathcal{I}$ によって構成される。本研究では、異なるドメイン $d_w, d_{w'}$ 間でユーザの重複が存在する ($\mathcal{U}^{d_w} \cap \mathcal{U}^{d_{w'}} \neq \emptyset$) シナリオを想定し、横断ユーザを通じた知識転移が可能である。一方で、アイテムは各ドメインに固有であると仮定し、異なるドメイン間で同一アイテムが共有されるケースは考慮しない。

3.5 推薦タスクの目的

本研究の目的は、入力として与えられた協調知識グラフ \mathcal{G}_{ckg} を用いて、各ドメイン $d_w \in \mathcal{D}$ におけるユーザ $u \in \mathcal{U}^{d_w}$ とアイテム $i \in \mathcal{I}^{d_w}$ の嗜好度を予測することである。具体的には、以下の予測関数 f を学習する。

$$\hat{y}_{ui}^{d_w} = f(u, i \mid \mathcal{G}_{\text{ckg}}, d_w) \quad (1)$$

ここで $\hat{y}_{ui}^{d_w}$ は、ドメイン d_w においてユーザ u がアイテム i とインタラクシヨンする確率 (嗜好スコア) を表す。

4 提案手法

本章では、負の転移の抑制と知識データの効果的な活用を目的とした、GNN ベースのフレームワーク DKGAT を詳述する。図 2 は、DKGAT のアーキテクチャ全体を示す。

4.1 表現の分離

DKGAT では、ユーザ・アイテム・知識データの埋め込み表現を、ドメイン間で共有されるドメイン不変表現 (inter) と、各ドメインに固有の固有表現 (intra) に分離して学習する。具体的には、各ユーザ $u \in \mathcal{U}$ 、アイテム $i \in \mathcal{I}$ 、知識エンティティ $e \in \mathcal{E} \setminus \mathcal{I}$ に対して、不変表現の埋め込みベクトル $\mathbf{x}_u^{\text{inter}}, \mathbf{x}_i^{\text{inter}}, \mathbf{x}_e^{\text{inter}} \in \mathbb{R}^H$ を定義する。また、ドメイン d_w に含まれるユーザ $u \in \mathcal{U}^{d_w}$ 、アイテム $i \in \mathcal{I}^{d_w}$ 、および当該ドメインで用いられる知識エンティティ e に対して、ドメイン固有表現の埋め込みベクトル $\mathbf{x}_{u,d_w}^{\text{intra}}, \mathbf{x}_{i,d_w}^{\text{intra}}, \mathbf{x}_{e,d_w}^{\text{intra}} \in \mathbb{R}^H$ を定義する。ここで、不変表現と固有表現の次元数はいずれも H とする。不変表現は全ドメインで共有して学習される一方で、各ドメインにおける予測ではそのドメインに含まれるユーザおよびアイテムに対応する不変表現のみを用いる。

4.2 知識グラフ向けアテンションエンコーダ

推薦に有用な知識の抽出とノイズ抑制のため、パラメータ化されたアテンション機構 [20], [26] を用いて、エンティティおよび関係に依存するコンテキストを適応的に捉える。知識グラフ \mathcal{G}_{kg} 上で、アイテム i に接続された知識エンティティ集合 \mathcal{N}_i からメッセージ集約を行う。具体的には、 $c = 1, \dots, C$ 層でアイテム埋め込みを以下のように更新する。

$$\mathbf{x}_i^{(c)} = \mathbf{x}_i^{(c-1)} + \sum_{e \in \mathcal{N}_i} \alpha(e, r_{e,i}, i) \mathbf{x}_e \quad (2)$$

ここで $\mathbf{x}_i^{(c)}$ は c 層目におけるアイテム i の埋め込み表現を表

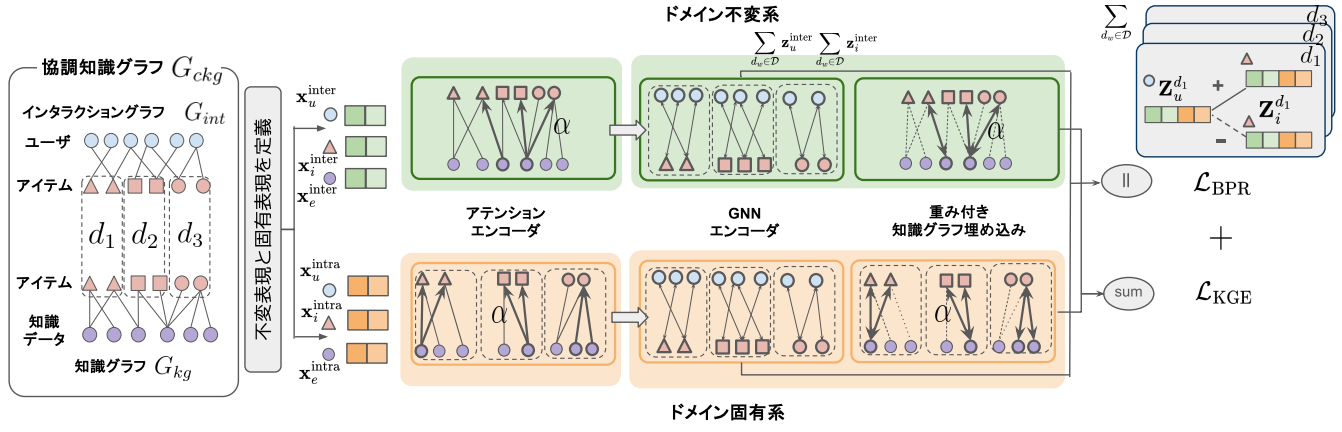


図 2: DKGAT の全体アーキテクチャ (ドメイン d_1, d_2, d_3 の例)

す。アテンション係数 $\alpha(e, r_{e,i}, i)$ は、関係 $r_{e,i}$ を介したアイテム i とエンティティ e の重要度を表し、以下で定義される。

$$s(e, r_{e,i}, i) = \text{LeakyReLU}(\mathbf{e}_{r_{e,i}}^\top \mathbf{W}[\mathbf{x}_e \parallel \mathbf{x}_i]),$$

$$\alpha(e, r_{e,i}, i) = \frac{\exp(s(e, r_{e,i}, i))}{\sum_{e' \in \mathcal{N}_i} \exp(s(e', r_{e',i}, i))} \quad (3)$$

ここで $\mathbf{x}_i, \mathbf{x}_e \in \mathbb{R}^H$ はそれぞれアイテムおよびエンティティの埋め込み、 $\mathbf{e}_{r_{e,i}} \in \mathbb{R}^H$ は関係 $r_{e,i}$ の埋め込み、 $\mathbf{W} \in \mathbb{R}^{H \times 2H}$ は学習可能な重み行列である。

本モデルでは、知識エンティティ埋め込み \mathbf{x}_e は層間で更新しない。これは、エンティティを安定した知識表現として保持し、ノイズの伝播を抑制するためである。一方で、アイテム埋め込みは各層においてアテンション機構による再重み付けが行われるため、有用な知識を段階的に強調する形で生成される。

本機構は不変系と固有系の二系統でそれぞれ独立に適用され、最終層 C におけるアイテム埋め込みを \mathbf{h}_i とする。これにより、不変アイテム埋め込み $\mathbf{h}_i^{\text{inter}}$ と固有アイテム埋め込み $\mathbf{h}_{i,d_w}^{\text{intra}}$ を生成する。知識エンティティ埋め込みについては、アテンションエンコーダに入力された埋め込みをそのまま \mathbf{h}_e とし、不変知識エンティティ埋め込み $\mathbf{h}_e^{\text{inter}}$ および固有知識エンティティ埋め込み $\mathbf{h}_{e,d_w}^{\text{intra}}$ とする。

4.3 インタラクショングラフ向け GNN エンコーダ

本研究の目的であるユーザの各アイテムに対する嗜好度を予測するために必要な、ユーザとアイテムの関係性を表現学習するため、GNN エンコーダを適用する。DKGAT ではインタラクショングラフ \mathcal{G}_{int} における各ドメイン d_w の二部グラフ $\mathcal{G}^{(d_w)}$ 上で、LightGCN 型の正規化近傍集約に residual を加えた伝播器を用いる [17], [32], [11]。各ドメインのノード集合 $V^{d_w} = U^{d_w} \cup I^{d_w}$ に対し、 $\ell = 1, \dots, L$ 層で以下の伝播を行う。

$$\mathbf{H}^{(\ell)} = (1 - \alpha_{\text{gnn}}) \tilde{\mathbf{A}}^{d_w} \mathbf{H}^{(\ell-1)} + \alpha_{\text{gnn}} \mathbf{H}^{(\ell-1)},$$

$$\tilde{\mathbf{A}}^{d_w} = (\mathbf{D}^{d_w})^{-1/2} \mathbf{A}^{d_w} (\mathbf{D}^{d_w})^{-1/2} \quad (4)$$

ここで $\mathbf{H}^{(\ell)} \in \mathbb{R}^{|V^{d_w}| \times H}$ は ℓ 層目における全ノードの潜在表現行列であり、各行が対応ノードの H 次元埋め込みベクトルを表す。初期表現行列 $\mathbf{H}^{(0)}$ は、ユーザ埋め込みとアイテム埋め込

みをノード順に連結することで構成される。ここで、アイテムは知識グラフ向けアテンションエンコーダによって得られた埋め込み $\mathbf{h}_i^{\text{inter}}$ および $\mathbf{h}_{i,d_w}^{\text{intra}}$ を初期値として用いる。ユーザについては知識グラフに含まれないため、 $\mathbf{x}_u^{\text{inter}}$ および $\mathbf{x}_u^{\text{intra}}$ は学習可能パラメータとして初期化する。 $\mathbf{A}^{d_w}, \mathbf{D}^{d_w}, \tilde{\mathbf{A}}^{d_w}$ は、それぞれグラフ $\mathcal{G}^{(d_w)}$ における隣接行列、次数行列、対称正規化隣接行列 [18], [19] である。 $\alpha_{\text{gnn}} \in [0, 1]$ は residual の重みである [3]。

DKGAT は不変系と固有系に独立した伝播器 F, F^{d_w} を持ち、各ドメイン d_w に対して

$$(\mathbf{z}_{u,d_w}^{\text{inter}}, \mathbf{z}_{i,d_w}^{\text{inter}}) = F(\mathcal{G}^{(d_w)}, \mathbf{x}_u^{\text{inter}}, \mathbf{h}_i^{\text{inter}}),$$

$$(\mathbf{z}_{u,d_w}^{\text{intra}}, \mathbf{z}_{i,d_w}^{\text{intra}}) = F^{d_w}(\mathcal{G}^{(d_w)}, \mathbf{x}_{u,d_w}^{\text{intra}}, \mathbf{h}_{i,d_w}^{\text{intra}}) \quad (5)$$

を得る。

さらに、ユーザおよびアイテムの不変表現は

$$\mathbf{z}_u^{\text{inter}} = \sum_{d_w \in \mathcal{D}_u} \mathbf{z}_{u,d_w}^{\text{inter}}, \quad \mathbf{z}_i^{\text{inter}} = \sum_{d_w \in \mathcal{D}_i} \mathbf{z}_{i,d_w}^{\text{inter}} \quad (6)$$

として定義する。ここで \mathcal{D}_u および \mathcal{D}_i は、それぞれユーザ u およびアイテム i が出現するドメイン集合を表す。本研究では、ユーザは複数ドメインにまたがって出現しうするため、ユーザの不変表現は当該ユーザが出現するドメインの不変表現のみを集約する。一方、本研究にて扱うグラフでは、アイテムは異なるドメイン間で重複が存在しないため、任意のアイテム i に対して $|\mathcal{D}_i| = 1$ が成り立つ。したがって、アイテムの不変表現は

$$\mathbf{z}_i^{\text{inter}} = \mathbf{z}_{i,d_w}^{\text{inter}} \quad (7)$$

と表される。本二系統 GNN により、ドメイン不変モデルではドメインを横断した潜在表現が得られ、ドメイン固有モデルでは各ドメイン固有の潜在表現が得られる。

ドメイン d_w におけるユーザ u およびアイテム i の最終的な潜在表現 $\mathbf{z}_{u,d_w} \in \mathbb{R}^{2H}$, $\mathbf{z}_{i,d_w} \in \mathbb{R}^{2H}$ は、不変表現と固有表現を結合して以下のように定義される。

$$\mathbf{z}_{u,d_w} = \mathbf{z}_u^{\text{inter}} \parallel \mathbf{z}_{u,d_w}^{\text{intra}},$$

$$\mathbf{z}_{i,d_w} = \mathbf{z}_i^{\text{inter}} \parallel \mathbf{z}_{i,d_w}^{\text{intra}} \quad (8)$$

ここで、 \parallel はベクトルの結合操作を表す。また、不変表現 $\mathbf{z}_u^{\text{inter}}$ および $\mathbf{z}_i^{\text{inter}}$ は、それぞれドメイン d_w に含まれるユーザ u およびアイテム i に対応する表現のみを用いられることに留意する。すなわち、不変表現は全ドメインで共有して学習されるが、各ドメインにおける最終表現の構成では、当該ドメインに含まれるノードに対応する不変表現とそのドメイン固有表現のみが結合される。

以上の GNN エンコーダによって得られた表現に基づき、内積によるスコア関数を用いて嗜好度を算出する。

$$\hat{y}_{ui}^{d_w} = f(u, i | d_w) = (\mathbf{z}_{u,d_w})^\top \mathbf{z}_{i,d_w} \quad (9)$$

学習には推薦手法 [5] [13] で広く用いられる Bayesian Personalized Ranking (BPR) [13] 損失を用い、各ドメインの損失の総和を最終的な推薦損失とする。

$$\mathcal{L}_{\text{BPR}} = \sum_{d_w \in \mathcal{D}} \sum_{(u,i,j) \in \mathcal{O}^{d_w}} -\ln \sigma(\hat{y}_{ui}^{d_w} - \hat{y}_{uj}^{d_w}) \quad (10)$$

ここで \mathcal{O}^{d_w} は、ドメイン d_w において観測されたアイテム i と未観測アイテム j の組からなる学習サンプル集合である。

4.4 重み付き知識グラフ埋め込み

抽出された知識を埋め込み空間における制約として明示的に反映させるため、知識グラフ埋め込みを適用する。DKGAT では TransE [2] に基づく知識グラフ埋め込みを採用する。知識グラフ \mathcal{G}_{kg} に含まれる各トリプレット $(e, r, i) \in \mathcal{T}$ に対し、以下の距離関数を定義する。

$$d(e, r, i) = \|\mathbf{h}_e + \mathbf{e}_{r_{e,i}} - \mathbf{h}_i\|_2^2 \quad (11)$$

ここで \mathbf{h}_e , \mathbf{h}_i , $\mathbf{e}_{r_{e,i}}$ は、それぞれ知識グラフ向けアテンションエンコーダによって得られたエンティティ e の埋め込み、アイテム i の埋め込み、関係 $r_{e,i}$ の埋め込みである。DKGAT では、知識グラフ埋め込みの学習において負例サンプリングを行い、マージンランキング損失に基づく学習を行う。具体的には、正例トリプレット (e, r, i) に対して、エンティティ e および関係 r を条件として負例アイテム i^- を $i^- \sim \mathcal{N}(e, r)$ から一様サンプリングし、負例トリプレット (e, r, i^-) を構成する。さらに、知識トリプレットの推薦への寄与度が一様ではない点に着目し、知識グラフ向けアテンションエンコーダによって得られたトリプレット重要度 $\alpha(e, r, i)$ を重みとして導入する。これにより、推薦にとって重要であると判断された知識構造に対して、より強い制約を課すことが可能となる。以上を踏まえ、重み付き知識グラフ埋め込み損失を以下のように定義する。

$$\ell = \sum_{(e,r,i) \in \mathcal{T}} \mathbb{E}_{i^- \sim \mathcal{N}(e,r)} \left[\alpha(e, r, i) \cdot \max(0, \gamma + d(e, r, i) - d(e, r, i^-)) \right] \quad (12)$$

ここで $\gamma > 0$ はマージンを表すハイパーパラメータである。直感的には、アテンション重み $\alpha(e, r, i)$ が大きい、すなわち推薦にとって重要であると判定された知識トリプレットほど、正例と負例の距離差に対する制約が強く課される。これにより、知

識グラフに含まれるノイズの影響を抑制しつつ、ドメイン横断およびドメイン固有の双方において有用な知識構造を優先的に学習することが可能となる。本重み付き知識グラフ埋め込みは、ドメイン不変表現およびドメイン固有表現それぞれに対して独立に適用し、ドメイン不変表現による損失 ℓ_{inter} とドメイン固有表現による各損失 $\ell_{\text{intra}}^{d_w}$ を用いて最終的な知識グラフ埋め込み損失を以下で定義する。

$$\mathcal{L}_{\text{KGE}} = \ell_{\text{inter}} + \sum_{d_w \in \mathcal{D}} \ell_{\text{intra}}^{d_w} \quad (13)$$

4.5 DKGAT の目的関数

最終的な目的関数 \mathcal{L} は、推薦損失、知識グラフ損失、および L2 正則化項を統合して以下のように定義される。

$$\mathcal{L} = \mathcal{L}_{\text{BPR}} + \lambda_{\text{KGE}} \mathcal{L}_{\text{KGE}} + \lambda_{\text{reg}} \|\Theta\|_2^2 \quad (14)$$

ここで λ_{KGE} および λ_{reg} はハイパーパラメータであり、 Θ は学習可能な全パラメータ集合を表す。

5 実 験

本章では、提案手法 DKGAT の有効性を検証するため、以下のリサーチクエスチョンに基づいて実験を行う。

- **RQ1:** DKGAT は従来の MDR 手法および KG ベース推薦手法と比較して推薦精度を向上できるか。
- **RQ2:** DKGAT を構成する各要素は推薦精度にどのように寄与するか。
- **RQ3:** DKGAT により学習された埋め込み表現は、ドメイン不変表現とドメイン固有表現という設計意図を反映しているか。
- **RQ4:** DKGAT は知識グラフから、不変表現および固有表現それぞれに有用な知識を選択的に抽出できているか。
- **RQ5:** DKGAT の主要ハイパーパラメータは性能にどのような影響を与えるか。

5.1 データセット

実験には、広範なドメインを含む Amazon Review Data (2023) [6] を使用した。本研究では、ドメインを横断して利用可能な知識データが存在し、かつ比較的インタラクション数の多い「Books」「Movies and TV (Movies)」「Electronics (Elec)」「Musical Instruments (Instr)」の 4 ドメインを選定した。また、計算効率の観点から本研究では、各ドメインについてユーザを 20% ランダムサンプリングしたデータを用いて評価した。表 1 にサンプリング後のインタラクションデータの統計量を示す。また、知識データとして、Books および Movies には「ジャンル」、Elec および Instr には「ブランド」を抽出して知識グラフを構築した。これにより、Books-Movies 間ではジャンル、Elec-Instr 間ではブランドを介した知識共有が可能となる。表 2 に知識データの統計量を示す。

さらに、実験に用いるマルチドメイングラフは、表 3 に示す 3 種類の構成を用いる。G2 は Books と Movies からなる構成であり、主としてジャンル知識を介した転移を評価する。G3 は

Elec と Instr からなる構成であり、主としてブランド知識を介した転移を評価する。G1 はこれら 4 ドメインすべてを統合した構成であり、横断ユーザおよび知識データの双方の観点から、より複雑なマルチドメイン環境における知識転移を評価する。ここでユーザ重複度とは、2 つのドメインに共通して出現するユーザの割合（各ドメインに含まれる全ユーザ数に対する比率）を表し、G1 については全ドメインペアで平均した値を示す。また知識データ重複度は、各知識データのうちドメインをまたいで共有される知識データの割合を表し、G1 については G2 および G3 の平均値を示す。

表 1: インタラクションデータの統計情報

ドメイン	#ユーザ	#アイテム	#インタラクション	密度
Books	155,120	424,960	1,916,815	2.89×10^{-5}
Movies	131,204	177,054	1,494,699	6.42×10^{-5}
Elec	328,556	326,943	3,097,586	2.88×10^{-5}
Instr	11,596	21,306	103,361	4.19×10^{-4}

表 2: 知識データの統計情報

ドメイン	知識データ	#アイテム	#知識データ	#トリプレット
Books	ジャンル	416,009	1,084	1,246,860
Movies	ジャンル	102,655	609	344,358
Elec	ブランド	267,875	40,720	267,875
Instr	ブランド	14,718	2,997	14,718

表 3: グラフデータにおける重複度情報

ドメインの組み合わせ	ユーザ重複度	知識データ重複度
G1: Books & Movies & Elec & Instr	0.0466	0.0255
G2: Books & Movies	0.0974	0.0297
G3: Elec & Instr	0.0196	0.0212

5.2 実験設定

5.2.1 評価手順と評価指標

各ドメインにおいて、各ユーザのレビューを時系列順に並べ、最後の 1 件を test、最後から 2 件目を validation、残りを train とする leave-one-out 評価を採用する。評価時には、各ユーザについて train および validation で未観測の全アイテムを候補集合とし、その上位 $K = 20$ 件に基づいて HR@20 および NDCG@20 を算出する。評価の信頼性を高めるため、各実験は乱数 seed を変えて 5 回繰り返し、その平均性能を報告する。表 4 および表 5 における統計的有意差は、5 回の実行結果に基づく最良ベースラインとの paired t-test により検定した。

5.2.2 実装の詳細

提案手法の実装には PyTorch および DGL [23] を用いた。GNN エンコーダの residual 重み α_{gnn} は 0.2、レイヤ数 L は 2 とした。知識グラフ向けアテンションエンコーダの層数 C は 2 とした。知識グラフ埋め込み損失の重み λ_{KGE} は 0.01、正則化項の重み λ_{reg} は 1×10^{-4} とした。BPR および知識グラフ埋め込みはミニバッチ学習とし、ミニバッチサイズはそれぞれ

4096, 8192 とした。埋め込み次元は 128 とし、不変表現および固有表現にはそれぞれ 64 次元を割り当てた。

5.3 全体的な推薦性能 (RQ1)

RQ1 を評価するために、以下のモデルを比較対象とする。

- **EDDA [11]**: ユーザ・アイテムについて、埋め込みレベルで不変・固有表現を分離する GNN ベースの MDR 手法。
- **FM [14] / NFM [4]**: 特徴量間の相互作用を捉える代表的なメタデータ活用手法。
- **KGAT [24]**: アテンション機構を用いて知識グラフ上の高次近傍情報を集約する KG ベース推薦手法。
- **KGCL [29]**: 対照学習を導入し、知識グラフのノイズ耐性とデータスパース性の緩和を図る KG ベース推薦手法。

なお、公平な比較のため、埋め込み次元、学習率、負例数、バッチサイズ、および早期終了条件を可能な限り統一し、各ベースラインは validation 指標に基づきチューニングを行った。

表 4 は、各グラフ構成における各モデルの推薦性能をドメインごとに示したものである。実験結果から、以下の知見が得られた。

- **全体性能**: DKGAT はすべてのグラフ構成 (G1/G2/G3) および評価指標 (HR@20/NDCG@20) において一貫して最良性能を達成した。
- **負の転移の抑制**: ドメインの区別を明示的に持たない FM/NFM や KGAT/KGCL と比較して、DKGAT は一貫して大きな改善を示した。特に、4 ドメインを統合した G1 において、ドメインの区別を持たない手法は精度劣化が見られる一方で、DKGAT は精度劣化を最小限に抑えつつ多くのドメインおよび指標において精度向上が見られる。これらより、不変表現と固有表現の分離により、複数ドメインからの信号を効果的に統合しつつ負の転移を抑制できていることが示唆される。
- **知識グラフ統合の有効性**: MDR 手法である EDDA と比較しても、DKGAT は全設定で優位であった。これは、ユーザ・アイテムの行動データに加えて、アイテム属性に基づく知識グラフを統合し、関係性を考慮した表現学習を行うことで、横断ユーザが少ない状況でも嗜好推定を補強できたことを示している。
- **グラフ構成の違いによる傾向**: G2 では Books と Movies の双方で大きな改善が見られ、ジャンル知識を介した転移が有効に機能していると考えられる。一方、G3 では Elec での改善幅は比較的小さいものの、Instr では大きな改善が確認された。これは、データ規模の小さいドメインほど、他ドメインおよび知識グラフからの補完的情報の恩恵を受けやすいことを示唆する。

5.4 アブレーション研究 (RQ2)

RQ2 を評価するために、提案モデルの各構成要素の寄与を検証する。各設定は以下の通りである。

- **w/o intra**: ドメイン固有表現を用いない設定 (不変表現の次元数は 128)。

表 4: 各グラフ構成 (G1/G2/G3) における推薦性能比較 (%) . 評価指標の HR と NDCG は上位 $K = 20$ 件 (HR@20, NDCG@20) で評価する. 太字は最良結果, 下線は 2 番目に良い結果を示す. * は最良ベースラインに対する統計的有意差 ($p < 0.05$) を示す. Impr.(%) は各ターゲットドメインにおける最良ベースラインに対する相対改善率を示す.

Method	G1: Books & Movies & Elec & Instr								G2: Books & Movies				G3: Elec & Instr			
	Books		Movies		Elec		Instr		Books		Movies		Elec		Instr	
	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG
EDDA	<u>1.95</u>	<u>0.79</u>	<u>3.57</u>	<u>1.46</u>	<u>2.00</u>	<u>0.82</u>	<u>5.77</u>	<u>2.39</u>	<u>1.93</u>	<u>0.80</u>	3.50	1.45	1.96	0.81	<u>5.12</u>	2.10
FM	1.28	0.49	2.47	1.07	1.08	0.31	3.44	1.31	1.33	0.51	2.61	1.13	1.14	0.39	3.57	1.47
NFM	1.37	0.56	2.66	1.18	1.12	0.35	3.71	1.44	1.43	0.58	2.83	1.23	1.22	0.43	3.86	1.59
KGAT	1.50	0.60	3.18	1.35	1.31	0.54	4.09	1.68	1.64	0.67	<u>3.56</u>	<u>1.46</u>	1.36	0.56	4.13	1.75
KGCL	1.57	0.61	2.99	1.24	1.93	0.79	5.01	2.12	1.73	0.68	3.24	1.29	<u>2.01</u>	<u>0.84</u>	5.11	<u>2.31</u>
DKGAT	2.85*	1.13*	4.69*	1.92*	2.30*	0.93*	6.68*	2.80*	2.88*	1.16*	4.67*	1.92*	2.14*	0.88*	6.18*	2.63*
Impr.(%)	+46.2	+43.0	+31.4	+31.5	+15.0	+13.4	+15.8	+17.2	+49.2	+45.0	+31.2	+31.5	+6.5	+4.8	+20.7	+13.9

表 5: DKGAT の各構成要素を除去した場合のアプレーション結果 (%) . 評価指標の HR と NDCG は上位 $K = 20$ 件 (HR@20, NDCG@20) で評価する. 太字は最良結果, 下線は 2 番目に良い結果を示す. * は最良ベースラインに対する統計的有意差 ($p < 0.05$) を示す.

Component	G1: Books & Movies & Elec & Instr								G2: Books & Movies				G3: Elec & Instr			
	Books		Movies		Elec		Instr		Books		Movies		Elec		Instr	
	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG
DKGAT (full)	2.85*	1.13*	4.69*	1.92*	2.30*	0.93*	6.68*	2.80*	2.88*	1.16*	4.67*	1.92*	<u>2.14</u>	0.88	6.18	2.63
w/o intra	2.15	0.87	4.01	1.66	2.11	0.87	5.81	2.44	2.18	0.90	4.08	1.71	2.07	0.84	<u>6.35</u>	2.82*
w/o inter	2.29	0.92	4.07	1.68	2.11	0.85	6.17	2.59	2.31	0.93	3.97	1.63	2.12	<u>0.87</u>	6.18	2.63
w/o KG	2.36	0.96	4.02	1.69	2.16	0.88	5.87	2.50	2.33	0.93	4.13	1.73	2.06	0.84	6.04	2.57
w/o att	2.34	0.95	4.12	1.70	<u>2.20</u>	0.89	6.29	2.62	2.44	0.98	4.22	1.76	2.15	0.88	5.89	2.47
w/o w-KGE	<u>2.60</u>	<u>1.05</u>	<u>4.45</u>	<u>1.83</u>	<u>2.20</u>	<u>0.90</u>	<u>6.39</u>	<u>2.67</u>	<u>2.69</u>	<u>1.10</u>	<u>4.51</u>	<u>1.81</u>	2.15	0.88	6.49*	<u>2.74</u>

- **w/o inter:** ドメイン不変表現を用いない設定 (固有表現の次元数は 128) .
- **w/o KG:** 知識グラフを用いず, インタラクショングラフのみを使う設定.
- **w/o att:** 知識グラフは用いるが, アテンションエンコーダを用いない設定. この場合, 近傍知識は一樣重みで集約される.
- **w/o w-KGE:** 知識グラフは用いるが, 重み付き知識グラフ埋め込みを用いない設定.

表 5 に各構成要素を除去した場合の推薦性能を示す. 実験結果から分かる各構成の効果を下に示す.

- **不変/固有表現の分離の効果:** w/o intra および w/o inter の双方で, 特に G1 (4 ドメイン統合) において明確な性能低下が確認された. これは, 多ドメイン統合環境では, 不変表現による共通嗜好の抽出と, 固有表現によるドメイン文脈の保持を両立することが性能安定化に寄与していることを示している. 一方で, G3-Instr のような小規模・少数ドメイン構成では, w/o intra が full を上回る指標も観測された. これは, 小規模ドメインでは固有表現を十分に学習するためのデータが不足し, 固有表現がノイズあるいは過学習要因として働く可能性を示唆する.
- **知識グラフ統合の効果:** w/o KG ではほぼすべての設定で

性能が低下しており, 知識データが補助情報として有効に機能していることが確認できる. 特に G1 のような多ドメイン構成では低下傾向がより明確であり, 知識グラフがドメイン間の知識転移を促進していると考えられる. また, 知識グラフは各ドメイン内でもアイテムの属性的・意味的關係を補完し, ユーザ嗜好の推定を精緻化していると考えられる.

- **アテンション機構の効果:** w/o att は一部の設定では full と近い性能を示すものの, 全体としては full より低い結果となった. 特に G1 や G2 のように知識トリプレットが多い構成では, アテンション導入による改善幅が大きい. これは, 関係数が増えるほど, 有用な知識を選択的に抽出する仕組みが重要になることを示唆する.
- **重み付き知識グラフ埋め込みの効果:** w/o w-KGE は多くの設定で 2 番目に良い結果を示し, G3 の一部指標では full を上回る結果も確認された. これは, 重要度に基づく強い制約が, 関係性の少ない小規模構成では過度な正則化として働く可能性を示唆する. 一方で, G1 や G2 のようにより多くの知識関係を含む構成では, full がより一貫して高い性能を示しており, 重み付き KGE がノイズ抑制と重要関係の強調の両面で寄与していると考えられる.
- **全体考察:** 各構成要素はすべての状況で常に最適というわ

けではないが、多ドメイン・多関係性の設定では full モデルが最も安定して高性能である。したがって、DKGAT の設計は、ドメイン数や関係性が増す複雑なマルチドメイン環境において特に有効であると考えられる。

5.5 埋め込み表現の可視化 (RQ3)

RQ3 では、提案手法が学習した埋め込み表現を可視化することによって、ドメイン不変表現とドメイン固有表現が設計意図を反映したものとなっているかを確認する。可視化には t-SNE [10] を用い、学習済み埋め込みベクトルを 2 次元空間に投影した。図 3 は、Books/Movies/Elec/Instr の 4 ドメインを含む G1 構成で学習したアイテム埋め込みの分布を示す。

可視化結果より、ドメイン固有表現はドメインごとに比較的に明確に分離したクラスターを形成しており、各ドメインの文脈に特化した特徴が捉えられていることが分かる。一方で、ドメイン不変表現は複数ドメインの分布がより大きく重なっており、各ドメインに共通する潜在的特徴を抽出できていることが示唆される。以上より、DKGAT は設計意図を反映できており、各ドメインの固有性を保持しつつ、ドメインをまたいで共有される特徴を表現する埋め込み学習を実現できていると考えられる。

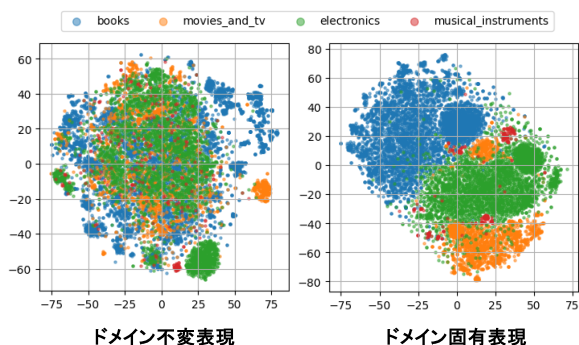


図 3: 学習済みアイテム埋め込みの t-SNE 可視化結果 (G1 構成)

5.6 重要知識データの例 (RQ4)

RQ4 では、提案手法の知識グラフアテンションエンコーダが、不変表現および固有表現それぞれに対して有用な知識を選択的に抽出できているかを定性的に確認する。具体的には、学習済みモデルが付与したアテンション重みを参照し、各アイテムに対する知識データの重要度を出力した。図 4 は、Books/Movies の 2 ドメインからなる G2 構成において、あるアイテムに対する「ジャンル」知識のアテンション重みの分布例を示す。

結果より、ドメイン不変空間では、Books ドメインのアイテムに対しては「Literature & Fiction」、Movies ドメインのアイテムに対しては「Movies & TV」など、ドメイン横断的な解釈に結びつきやすい属性（文学やフィクション好き、映画やテレビ作品好き）に高い重要度が付与されていることが確認できる。一方で、ドメイン固有空間では、同一アイテムに対する重みが相対的に分散しており、特定の属性に過度に依存せず、より多面的な属性情報を取り込んでいる様子が見られた。これらの結

果は、不変空間ではドメイン共通の嗜好推定に有用な知識が強調され、固有空間では各ドメイン文脈に依存する多面的な知識が利用されている可能性を示唆している。

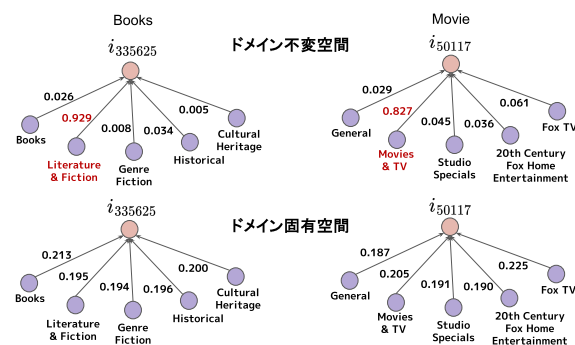


図 4: 知識グラフアテンションによる重要度の例. あるアイテムに対して学習されたアテンション重みの分布を示す

5.7 パラメータ分析 (RQ5)

RQ5 では、DKGAT の主要ハイパーパラメータが性能に与える影響を分析する。具体的には、(i) 知識グラフ埋め込み損失の重み λ_{KGE} , (ii) GNN 伝播における residual の重み α_{gnn} , (iii) GNN 層数 L を対象とし、他の設定を固定したうえで各値を変化させて評価した。図 5 に各グラフ構成 (G1/G2/G3) での各ドメインの平均値を示す。得られた知見は以下の通りである。

- λ_{KGE} の影響:** 本実験では、 λ_{KGE} に対して性能は過度には敏感ではなく、各グラフ構成において比較的一定の精度を示した。グラフ構成によりややばらつきが見られるが、これはドメイン数や知識トリプレット数の違いにより、知識制約が表現学習に与える影響度が異なるためと考えられる。特に G3 のように小規模・少数ドメインを含むグラフでは、RQ2 の結果と同様に、知識グラフ埋め込みの制約が相対的に強く働き、過度な正則化となる可能性が示唆される。
- α_{gnn} の影響:** 本実験では、 α_{gnn} を大きくしすぎると性能が低下する傾向が見られた。これは自己情報への依存が強まり、グラフ構造を通じた高次関係の伝播が不十分になるためと解釈できる。一方で、G2 は $\alpha_{gnn} = 0.3$ において最も高い性能を示しており、グラフ構成によっては残差接続が過度な平滑化を抑制し、自己情報と近傍情報のバランスを取る上で有効に働くことが示唆される。したがって、 α_{gnn} は一貫して小さいほど良いのではなく、近傍集約を十分に保ちつつ過平滑化を緩和できる範囲で設定することが重要である。
- 層数 L の影響:** 本実験では計算コストの観点から、層数 L を 1, 2, 3 の場合で比較し、すべてのグラフ構成において $L = 2$ が最良の性能を示した。本研究の GNN エンコーダはインタラクショングラフ（ユーザー・アイテムからなる二部グラフ）上で伝播を行うため、同種ノード間の文脈を間接的に取り込むことが可能な偶数ホップである $L = 2$ が有効に機能したと考えられる。また別の視点では、 $L = 1$ では高次近傍情報が不足しやすく、ドメイン横断的な知識

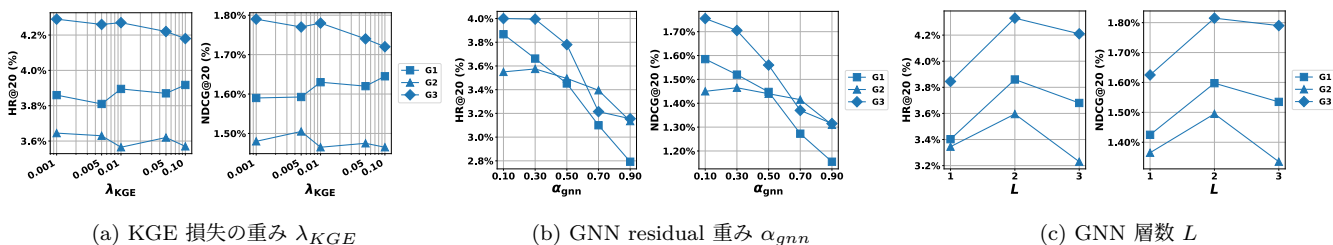


図 5: 主要ハイパーパラメータに対する性能変化

伝播が十分でない可能性が考えられ、 $L = 3$ では過平滑化やノイズ伝播の影響により性能が低下する可能性も考えられる。

6 結 論

本研究では、マルチドメイン推薦 (MDR) における負の転移の抑制と、外部知識の効果的な活用を両立するため、協調知識グラフ上でドメイン不変表現とドメイン固有表現を分離して学習する GNN ベース推薦手法 DKGAT を提案した。提案手法では、ユーザ・アイテム・知識データについてそれぞれ不変表現と固有表現を定義し、それぞれに対して独立した GNN エンコーダを適用する二系統の表現学習フレームワークを構築することで、ドメイン間で共有される嗜好とドメイン固有の嗜好を同時に捉えることを可能にした。さらに、知識グラフ上のアテンション機構および重み付き知識グラフ埋め込みを導入することで、推薦に有用な知識関係を選択的に強調しながら表現学習を行い、外部知識を効果的に推薦モデルへ取り込むことを可能にした。

Amazon Review Data を用いた複数ドメイン設定での実験の結果、DKGAT は既存の MDR 手法および KG ベース推薦手法と比較して一貫して高い推薦性能を示した。また、アブレーション研究から、不変/固有表現の分離、知識グラフ統合、アテンション機構、および重み付き知識グラフ埋め込みが相補的に機能していることを確認した。さらに、埋め込み表現の可視化およびアテンション重みの分析により、提案手法がドメイン共通の特徴とドメイン固有の特徴を異なる表現空間で捉え、それぞれに対して有用な知識を選択的に利用していることが示唆された。

今後の展望としては、より多様な知識関係を含む知識グラフへの拡張や、ドメイン数がさらに多い大規模設定においてもドメイン間の相関関係や多様な属性情報を適応的に学習できるような、スケーラブルなアーキテクチャへの発展を目指す予定である。

文 献

- [1] Alejandro Ariza-Casabona, Bartłomiej Twardowski, and Tri Kurniawan Wijaya. Exploiting graph structured cross-domain representation for multi-domain recommendation. In *European conference on information retrieval*, pp. 49–65. Springer, 2023.
- [2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, Vol. 26, , 2013.
- [3] Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized pagerank. *arXiv preprint arXiv:1810.05997*, 2018.
- [4] Xiangnan He and Tat-Seng Chua. Neural factorization machines for sparse predictive analytics. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 355–364, 2017.
- [5] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, pp. 639–648, 2020.
- [6] Yupeng Hou, Jiacheng Li, Zhankui He, An Yan, Xiusi Chen, and Julian McAuley. Bridging language and items for retrieval and recommendation. *arXiv preprint arXiv:2403.03952*, 2024.
- [7] Yuchen Jiang, Qi Li, Han Zhu, Jinbei Yu, Jin Li, Ziru Xu, Huihui Dong, and Bo Zheng. Adaptive domain interest network for multi-domain recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 3212–3221, 2022.
- [8] Bin Li, Qiang Yang, and Xiangyang Xue. Transfer learning for collaborative filtering via a rating-matrix generative model. In *Proceedings of the 26th annual international conference on machine learning*, pp. 617–624, 2009.
- [9] Shengchao Liu, Yingyu Liang, and Anthony Gitter. Loss-balanced task weighting to reduce negative transfer in multi-task learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33, pp. 9977–9978, 2019.
- [10] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, Vol. 9, No. Nov, pp. 2579–2605, 2008.
- [11] Wentao Ning, Xiao Yan, Weiwen Liu, Reynold Cheng, Rui Zhang, and Bo Tang. Multi-domain recommendation with embedding disentangling and domain alignment. In *Proceedings of the 32nd ACM international conference on information and knowledge management*, pp. 1917–1927, 2023.
- [12] Xichuan Niu, Bofang Li, Chenliang Li, Jun Tan, Rong Xiao, and Hongbo Deng. Heterogeneous graph augmented multi-scenario sharing recommendation with tree-guided expert networks. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pp. 1038–1046, 2021.
- [13] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618*, 2012.
- [14] Steffen Rendle, Zeno Gantner, Christoph Freudenthaler, and Lars Schmidt-Thieme. Fast context-aware recommendations with factorization machines. In *Proceedings of the*

- 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 635–644, 2011.
- [15] Abdulaziz Samra, Evgeny Frolov, Alexey Vasilev, Alexander Grigorevskiy, and Anton Vakhrushev. Cross-domain latent factors sharing via implicit matrix factorization. In *Proceedings of the 18th ACM conference on recommender systems*, pp. 309–317, 2024.
- [16] Xiang-Rong Sheng, Liqin Zhao, Guorui Zhou, Xinyao Ding, Binding Dai, Qiang Luo, Siran Yang, Jingshan Lv, Chi Zhang, Hongbo Deng, et al. One model to serve all: Star topology adaptive recommender for multi-domain ctr prediction. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 4104–4113, 2021.
- [17] Yixin Su, Rui Zhang, Sarah Erfani, and Zhenghua Xu. Detecting beneficial feature interactions for recommender systems. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35, pp. 4357–4365, 2021.
- [18] Yixin Su, Rui Zhang, Sarah M. Erfani, and Junhao Gan. Neural graph matching based collaborative filtering. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pp. 849–858, 2021.
- [19] Yixin Su, Yunxiang Zhao, Sarah Erfani, Junhao Gan, and Rui Zhang. Detecting arbitrary order beneficial feature interactions for recommender systems. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 1676–1686, 2022.
- [20] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. *International Conference on Learning Representations*, 2018. accepted as poster.
- [21] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. Knowledge graph convolutional networks for recommender systems. In *The world wide web conference*, pp. 3307–3313, 2019.
- [22] Liyuan Wang, Mingtian Zhang, Zhongfan Jia, Qian Li, Chenglong Bao, Kaisheng Ma, Jun Zhu, and Yi Zhong. Afec: Active forgetting of negative transfer in continual learning. *Advances in Neural Information Processing Systems*, Vol. 34, pp. 22379–22391, 2021.
- [23] Minjie Wang, Da Zheng, Zihao Ye, Quan Gan, Mufei Li, Xiang Song, Jinjing Zhou, Chao Ma, Lingfan Yu, Yu Gai, Tianjun Xiao, Tong He, George Karypis, Jinyang Li, and Zheng Zhang. Deep graph library: A graph-centric, highly-performant package for graph neural networks. *arXiv preprint arXiv:1909.01315*, 2019.
- [24] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 950–958, 2019.
- [25] Xiang Wang, Tinglin Huang, Dingxian Wang, Yancheng Yuan, Zhenguang Liu, Xiangnan He, and Tat-Seng Chua. Learning intents behind interactions with knowledge graph for recommendation. In *Proceedings of the web conference 2021*, pp. 878–887, 2021.
- [26] Lianghao Xia, Chao Huang, Yong Xu, Peng Dai, Xiyue Zhang, Hongsheng Yang, Jian Pei, and Liefeng Bo. Knowledge-enhanced hierarchical graph transformer network for multi-behavior recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35, pp. 4486–4493, 2021.
- [27] Zixuan Xu, Penghui Wei, Shaoguo Liu, Weimin Zhang, Liang Wang, and Bo Zheng. Correlative preference transfer with hierarchical hypergraph network for multi-domain recommendation. In *Proceedings of the ACM web conference 2023*, pp. 983–991, 2023.
- [28] Yuhao Yang, Chao Huang, Lianghao Xia, and Chunzhen Huang. Knowledge graph self-supervised rationalization for recommendation. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 3046–3056, 2023.
- [29] Yuhao Yang, Chao Huang, Lianghao Xia, and Chenliang Li. Knowledge graph contrastive learning for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pp. 1434–1443, 2022.
- [30] Donghan Yu, Yiming Yang, Ruohong Zhang, and Yuexin Wu. Knowledge embedding based graph convolutional network. In *Proceedings of the web conference 2021*, pp. 1619–1628, 2021.
- [31] Fan Zhang, Qiuying Peng, Yulin Wu, Zheng Pan, Rong Zeng, Da Lin, and Yue Qi. Multi-graph based multi-scenario recommendation in large-scale online video services. In *Companion Proceedings of the Web Conference 2022*, pp. 1167–1175, 2022.
- [32] Rui Zhang, Bayu Distiawan Trisedya, Miao Li, Yong Jiang, and Jianzhong Qi. A benchmark and comprehensive survey on knowledge graph entity alignment via representation learning. *The VLDB Journal*, Vol. 31, No. 5, pp. 1143–1168, 2022.
- [33] Xinjun Zhu, Yuntao Du, Yuren Mao, Lu Chen, Yujia Hu, and Yunjun Gao. Knowledge-refined denoising network for robust recommendation. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval*, pp. 362–371, 2023.

グラフ構造を考慮した負例の信頼度推定に基づく推薦モデルの提案

曾根 良行[†] 道瀬 悠磨[†] 原 隆浩[†]

[†] 大阪大学 情報科学研究科 〒 565-0871 大阪府吹田市山田丘

E-mail: †{sone.yoshiyuki,douse.yuma,hara}@ist.osaka-u.ac.jp

あらまし 近年、正と負のフィードバックを統合的に扱う符号付きグラフベースの推薦モデルが注目を集めている。しかし、既存手法は観測されたすべての負のフィードバックを一律に絶対的な拒絶として扱う。実際の負のフィードバックには明確な嫌悪から誤操作などの曖昧な拒絶まで幅広い信頼性の違いが存在するため、これらを同等に扱うことはユーザの嗜好空間を著しく歪める原因となる。この課題に対処するため、本研究では負のフィードバックの信頼性を適応的に捉える符号付き推薦モデル SANE (Structure-Aware Negative Reliability Estimation for Signed Recommendation) を提案する。SANE は、符号付きグラフの階層的な構造情報から導出される構造的類似度を用いて各負のインタラクションの信頼性を動的に評価する。本手法はすべての負のエッジを一律に扱うのではなく、推定された信頼度に基づいて不確実な負のエッジへの重み付け学習を行う。そして、構造的に矛盾する信頼性の低い負のエッジに対しては、埋め込み表現の歪みを防ぐためにグラフ構造自体を動的に修正する。さらに、修正済みグラフを用いた対照学習を導入し、ポジティブな嗜好と構造的に裏付けられた真の嫌悪との間に明示的な反発力を働かせ、表現の識別能力を大幅に向上させる。実世界データセットを用いた実験の結果、SANE は従来手法に対し Recall@20 で最大 6.34% の性能向上を達成した。

キーワード 推薦システム, 符号認識型推薦, 負のフィードバック, 偽陰性検出

1 序 論

推薦システムは現代の Web サービスや電子商取引プラットフォームにおいて不可欠な基盤技術となっている。個々のユーザの嗜好に合わせた関連コンテンツを提供することで情報過多を軽減する。特に協調フィルタリングにおいては、graph neural network [9, 11, 30] や graph Transformer [5, 17, 28, 33] などのグラフ表現学習の発展が著しい進歩をもたらしている。これらの手法は、ユーザとアイテムのインタラクションを二部グラフとしてモデル化することで、グラフ構造に内在する高次の協調シグナルを効果的に捉え、最先端の推薦性能を達成している。しかし、これらの手法はクリックや購入などの正のフィードバックのみを活用しており、低評価やスキップといった負のフィードバックを活用していない。負のフィードバックはユーザが何を嫌うかという有用な手がかりを提供するため、正のデータだけでは捉えきれない多角的で包括的な嗜好モデリングを可能にする。[15, 23, 24, 27].

近年、この負のフィードバックの重要性に着目し、正負のインタラクションを符号付きグラフ (Signed graph) として統合的に学習する符号認識型の推薦モデルが登場している [4, 20, 26]. これらは従来の正例のみの手法を上回る性能を達成しているが、既存手法の多くは、観測された全ての負のインタラクションを一律に真の負例、すなわちユーザの明確な拒絶であると仮定して学習を行ってしまう [5, 14, 20]. しかし、現実のユーザ行動において観測される負例は一律ではなく、「明確な嫌悪」から、誤操作や文脈の不一致による「曖昧な拒絶」まで、幅広いスペクトルが存在する [15, 23]. 既存手法はこれらのすべての負例

を一律に絶対的な拒絶として厳密に扱うため、このような行動の複雑さを捉えられていない [2, 3]. 信頼性の低い曖昧な拒絶を明確な拒絶と同等に扱うことは、本来嗜好の近いユーザとアイテムの埋め込み表現を不当に遠ざけ、ユーザの嗜好空間の歪みを通じて推薦性能を著しく低下させる。

この課題に対処するため、本研究では負のインタラクションの信頼性を適応的に捉える符号付きグラフ推薦モデル SANE (Structure-Aware Negative Reliability Estimation for Signed Recommendation) を提案する。SANE は符号付きグラフの階層的な構造情報から導出される構造的類似度を用いて、各負のインタラクションの信頼性を動的に評価する。すべての負のエッジを一律に扱うのではなく、不確実な負のエッジに対して信頼度に基づく重み付けを適用し、推定された強度に基づいて学習への影響を調整する。さらに、構造的に矛盾する極めて信頼性の低い負のエッジを動的に修正し、埋め込み表現の歪みを防ぐ。重要な点は、この適応的な調整がユーザの嫌悪を無視するわけではなく、真の拒絶を正確に反映するように負の信号を較正していることである。さらに、修正された信頼度の高い符号付きグラフを活用し、修正済みグラフ対照学習モジュールを導入する。本モジュールは、埋め込み空間においてポジティブな嗜好と構造的に裏付けられた真の嫌悪との間に明示的な反発力を働かせ、学習表現の正負の識別能力を向上させる。

本研究の主な貢献は以下の通りである。

- 負のフィードバックにおける幅広い信頼度スペクトルに着目し、動的な信頼度推定と信頼度に基づく重み付けを利用する新たな符号付き推薦モデル SANE を提案する。
- 埋め込み表現における正負の識別能力を高めるため、ポジティブな嗜好と構造的に裏付けられた真の嫌悪を明示的に

対比させる修正済みグラフ対照学習モジュールを設計する。

- 3つの実世界データセットを用いた包括的な評価実験により、提案手法が既存の最先端手法と比較して有意な性能向上を達成することを示し、その有効性を実証する。

2 事前準備

本節では、主要な記法を定義し、符号付き推薦タスクの定式化を行う。

2.1 記法と定義

M 人のユーザからなる集合を $\mathcal{U} = \{u_1, u_2, \dots, u_M\}$, N 個のアイテムからなる集合を $\mathcal{I} = \{i_1, i_2, \dots, i_N\}$ とする。本研究では、ユーザとアイテム間のインタラクションを符号付き二部グラフ $\mathcal{G} = (\mathcal{U}, \mathcal{I}, \mathcal{E})$ としてモデル化する。エッジ集合 $\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^-$ は、互いに排反な2つの部分集合、すなわち正のエッジ集合 \mathcal{E}^+ と負のエッジ集合 \mathcal{E}^- から構成され、 $\mathcal{E}^+ \cap \mathcal{E}^- = \emptyset$ を満たす。ここで、 \mathcal{E}^+ はユーザ u がアイテム i に対して購入、高評価といった正のインタラクションを表し、 \mathcal{E}^- は低評価、スキップといった明示的な拒絶などの負の反応を表す。グラフ構造を代数的に表現するため、符号付きインタラクション行列 $\mathbf{A} \in \{0, 1, -1\}^{M \times N}$ を定義する。ここで、各要素 A_{ui} は次のように与えられる

$$A_{ui} = \begin{cases} 1 & \text{if } (u, i) \in \mathcal{E}^+ \\ -1 & \text{if } (u, i) \in \mathcal{E}^- \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

ここで、0はユーザ u とアイテム i の間にインタラクションが観測されていないことを示す。

2.2 問題設定

観測された符号付き二部グラフ \mathcal{G} が与えられたとき、本研究の目的は、ユーザ u の未観測アイテム i に対する嗜好スコア \hat{y}_{ui} を予測可能な推薦モデルを学習することである。この目的を達成するため、各ユーザ u とアイテム i をそれぞれ d 次元の埋め込みベクトル $\mathbf{e}_u, \mathbf{e}_i \in \mathbb{R}^d$ へとマッピングする。次に、これらの表現をスコアリング関数 $s(\mathbf{e}_u, \mathbf{e}_i)$ に入力し、ユーザの嗜好を推定する。最終的に、モデルは予測スコアに基づいて未観測アイテムをランク付けし、各ユーザに対して上位 K 件の推薦リストを生成する。

3 提案手法

本節では、負のインタラクションの信頼性を動的に推定することでユーザの嗜好をロバストに学習する、新たな符号付き推薦モデル SANE を提案する。図1に SANE の概要を示す。SANE は、以下の3つのモジュールから構成される。1つ目は高次の協調シグナルを捉える符号付きグラフエンコーダ (Signed Graph Encoder), 2つ目はインタラクションの信頼性を動的に定量化し偽陰性を修正する構造認識型ネガティブ信頼度推定 (Structure-Aware Negative Reliability Estimation)

モジュール、そして3つ目はインタラクションの信頼度に基づいて学習シグナルを調整し、ユーザの嗜好と嫌悪を対照的にモデリングすることで学習プロセスを洗練させる構造認識型最適化 (Structure-Aware Optimization) モジュールである。

3.1 符号付きグラフエンコーダ

符号付きグラフに内在する複雑な協調シグナルを効果的に捉えるため、バックボーンエンコーダとして Graph Transformer アーキテクチャを採用する。具体的には、SIGformer [5] で導入されたメカニズムを活用し、スペクトル情報とパス情報を Transformer に統合することで、符号付きグラフにおける大域のおよび局所的な構造パターンの両方を捉える。

3.1.1 Graph Transformer 層

エンコーダは複数の Transformer 層から構成される。初期 ID 埋め込み表現を $\mathbf{E}^{(0)}$ としたとき、 l 層目のノード埋め込み行列を $\mathbf{E}^{(l)} \in \mathbb{R}^{(M+N) \times d}$ と定義し、 l 層目におけるノード表現の集約プロセスは次のように定式化される

$$\mathbf{E}^{(l)} = \frac{1}{2} \left(\text{softmax} \left(\frac{\mathbf{Q}^{(l)}(\mathbf{K}^{(l)})^\top}{\sqrt{d}} + \mathbf{P}_s^{(l)} \right) + \text{softmax}(\mathbf{P}_p^{(l)}) \right) \mathbf{V}^{(l)} \quad (2)$$

ここで、 $\mathbf{Q}^{(l)} = \mathbf{K}^{(l)} = \mathbf{V}^{(l)} = \mathbf{E}^{(l-1)}$ はそれぞれ Query, Key, Value 行列を示す。また、 $\mathbf{P}_s^{(l)}$ と $\mathbf{P}_p^{(l)}$ はそれぞれ符号付きスペクトルエンコーディング (Signed Spectral Encoding: SSE) と符号付きパスエンコーディング (Signed Path Encoding: SPE) を表すバイアス項であり [6], これらのエンコーディングを導入することで、異なるスケールにおける大域的なスペクトル特性と局所的なパスに基づく協調関係をモデルに組み込んでいる。

3.2 構造認識型ネガティブ信頼度推定

符号付きグラフエンコーダから得られたノード表現に基づき、本モジュールはインタラクションの信頼性を定量化する構造的類似度指標を導入することで、信頼性の低いインタラクションの影響を軽減しグラフ構造を動的に洗練させる。

3.2.1 層間構造類似度

多くの既存研究におけるインタラクションの信頼度推定は、学習中の最終的な埋め込み表現間の類似度に基づいて行われる。しかし、これらの手法は自己強化的な循環問題を孕んでいる。すなわち、不確実なインタラクションを含んだまま学習された埋め込み表現は、真の嗜好構造を正確に反映できていない可能性が高い。そのような歪んだ表現空間上から得られる埋め込み表現を用いて再度信頼度を推定すると、本来は信頼性の低いインタラクションを誤って高い信頼度と評価してしまうリスクが生じる [32]。この誤った推定に基づいてモデルの最適化を進めることは、初期の推定誤差を学習過程で増幅・固定化させる結果となり、局所解への収束や学習の不安定化を招く要因となる。特に、正のインタラクションに比べて学習が困難な負のインタラクションにおいて、この傾向は顕著である。

この問題に対処するため、本研究では符号付きグラフエンコーダの0層目および1層目の出力を利用した構造的類似性に

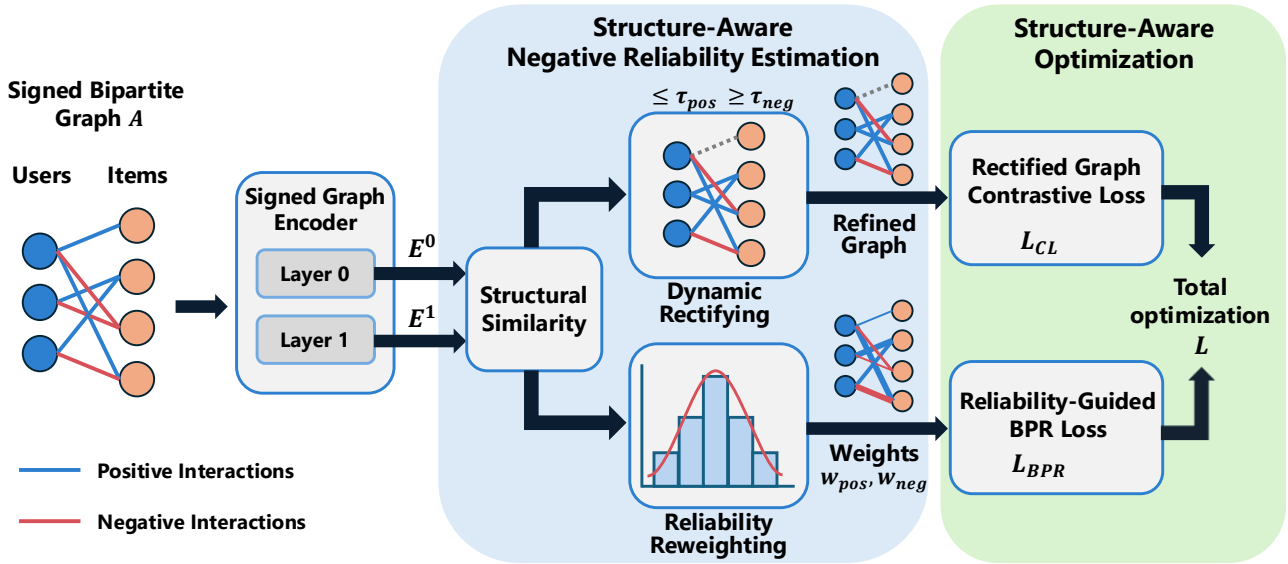


図 1: SANE の全体アーキテクチャ.

基づく新たな信頼度推定手法を提案する。この推定手法は、符号付きグラフエンコーダにおいて 0 層目 $\mathbf{E}^{(0)}$ がノード固有の特徴を表し、1 層目 $\mathbf{E}^{(1)}$ は符号付きグラフの近傍ノードから集約された局所的な構造コンテキストを捉えるという特性を利用する。ホモフィリー仮定 [12] に基づくと、ユーザー u とアイテム i の間に強い嗜好関係が存在する場合、それぞれのノード固有表現は互いの構造コンテキストと高い一致度を示すと考えられる、そのため、この層間の類似性を測ることでインタラクションの信頼性を評価する。これを定量化するため、まず初期埋め込み $\mathbf{E}^{(0)}$ と 1 層目の出力 $\mathbf{E}^{(1)}$ から抽出されたユーザーおよびアイテムベクトルに対して、 L_2 正規化を適用する。

$$\mathbf{h}_u^{(l)} = \frac{\mathbf{e}_u^{(l)}}{\|\mathbf{e}_u^{(l)}\|_2}, \quad \mathbf{h}_i^{(l)} = \frac{\mathbf{e}_i^{(l)}}{\|\mathbf{e}_i^{(l)}\|_2} \quad \text{for } l \in \{0, 1\}. \quad (3)$$

次に、ここで得られた正規化ベクトルを用い、2つの層を相互に参照することで生の構造的類似度 $s_{u,i}$ を計算する。

$$s_{u,i} = \frac{1}{2} \left((\mathbf{h}_u^{(0)} \cdot \mathbf{h}_i^{(1)}) + (\mathbf{h}_u^{(1)} \cdot \mathbf{h}_i^{(0)}) \right). \quad (4)$$

そして、この構造的類似度を確率的な信頼度スコアとして扱うため、得られたコサイン類似度を $[-1, 1]$ から $[0, 1]$ へと線形変換する。

$$\tilde{s}_{u,i} = \frac{s_{u,i} + 1}{2}. \quad (5)$$

この正規化された構造的類似度 $\tilde{s}_{u,i}$ は、対象のインタラクションの符号付きグラフにおける構造の整合性を評価するものであり、単一の層に基づく単純な類似度と比較して、埋め込み表現の歪みの影響を受けにくい堅牢な信頼度推定を可能にする。

3.2.2 動的符号付きグラフ修正

理想的には、観測されたインタラクションはユーザーの潜在的な嗜好構造と一致しているべきである。しかし、偽陰性インタラクションの存在はグラフに構造的な矛盾を引き起こす。すなわち、高い構造的整合性を持つにもかかわらず、エッジが負としてラベル付けされてしまう状態である。この矛盾を解消する

ため、本手法では潜在的な偽陰性を正のシグナルへと動的に修正するプロセスを導入する。具体的には、 \mathcal{E}^- 内の偽陰性を正のシグナルへと修正するための修正率を $\rho_{neg} \in [0, 1]$ 、 \mathcal{E}^+ から偽陽性を除去するためのフィルタリング率を $\rho_{pos} \in [0, 1]$ と定義する。これらの操作における適応的な閾値を決定するため、類似度スコアの集合 \mathcal{S} と確率 p に対する分位点関数 $Q(\mathcal{S}, p)$ を導入し、以下のように閾値を計算する。

$$\tau_{pos} = Q(\{\tilde{s}_{u,i} \mid (u,i) \in \mathcal{E}^+\}, \rho_{pos}) \quad (6)$$

$$\tau_{neg} = Q(\{\tilde{s}_{u,i} \mid (u,i) \in \mathcal{E}^-\}, 1 - \rho_{neg}). \quad (7)$$

これらの閾値を用いて、インタラクション行列 \mathbf{A}' を次のように更新する。

$$A'_{ui} = \begin{cases} 0 & \text{if } (u,i) \in \mathcal{E}^+ \wedge \tilde{s}_{u,i} < \tau_{pos} \\ 1 & \text{if } (u,i) \in \mathcal{E}^- \wedge \tilde{s}_{u,i} > \tau_{neg} \\ A_{ui} & \text{otherwise} \end{cases} \quad (8)$$

ここで、偽陽性は負のインタラクションである -1 ではなく、未観測を表す 0 へと修正される点に留意されたい。これは、誤クリックなどのインタラクションが明示的な嫌悪ではなく、単なる信頼性の低いシグナルであることを反映するためである。この一連のプロセスにより、学習されたユーザーの嗜好と整合するようにグラフ構造が能動的に再構築される。

3.3 構造認識型最適化

本節では、信頼度に基づく BPR 重み付けと修正済みグラフ対照学習を統合した構造認識型の最適化目的関数を提案する。その目的は、構造的な信頼度に基づいて学習シグナルの強度を適応的に調整するとともに、ユーザーの真の嗜好と信頼性の高い嫌悪とを明示的に対比させることである。

3.3.1 信頼度に基づく BPR 重み付け

符号付き推薦モデルにおける標準的な BPR 損失 [25] は、観測されたすべての負のフィードバックを同等に扱うため、現実

世界の負のフィードバックが本来持つ不均一性や信頼度の違いを見落としてしまう。例えば、一時的な文脈の不一致による拒絶や誤操作といった潜在的な偽陰性に対して、明示的で強い拒絶を表す真の負例と同等のパナルティを課すことは適切ではない [8].

このフィードバックの不均一性に対処するため、本手法では信頼度に基づく重み付き BPR 損失を以下のように提案する。

$$\begin{aligned} \mathcal{L}_{bpr} = & - \sum_{(u,i) \in \mathcal{E}^+} w_{pos}(u,i) \ln \sigma(\hat{y}_{ui} - \hat{y}_{uj}) \\ & - \sum_{(u,k) \in \mathcal{E}^-} w_{neg}(u,k) \ln \sigma(\hat{y}_{uj} - \hat{y}_{uk}), \end{aligned} \quad (9)$$

ここで、 $\hat{y}_{ui} = \mathbf{e}_u^\top \mathbf{e}_i$ は予測された嗜好スコアを示す。最終的なユーザおよびアイテム表現はすべての層の出力を平均化することで得られ、それぞれ $\mathbf{e}_u = \frac{1}{L+1} \sum_{l=0}^L \mathbf{e}_u^{(l)}$ および $\mathbf{e}_i = \frac{1}{L+1} \sum_{l=0}^L \mathbf{e}_i^{(l)}$ として定式化される。さらに、 j はユーザ u に対して一様にサンプリングされた未観測アイテムを表す。重み $w_{pos}(u,i)$ は、確信度の高い嗜好の学習を強化するために次のように定義する。

$$w_{pos}(u,i) = \frac{\tilde{s}_{u,i}}{\max_{(u',i') \in \mathcal{E}^+} \tilde{s}_{u',i'}}. \quad (10)$$

一方、負のインタラクションに対する重みは、嫌悪の強度を反映するように定義する。低い構造的類似度スコア $\tilde{s}_{u,k}$ は強度の高い真の負例を反映する強い構造的矛盾のシグナルと解釈し、対照的に高い $\tilde{s}_{u,k}$ は強度の弱い潜在的な偽陰性を意味すると解釈する。したがって、 $w_{neg}(u,k)$ は次のように定義する。

$$w_{neg}(u,k) = \beta \cdot \frac{1 - \tilde{s}_{u,k}}{\max_{(u',k') \in \mathcal{E}^-} (1 - \tilde{s}_{u',k'})}. \quad (11)$$

この定式化により、モデルは不確実な負例サンプルからの勾配の影響を軽減しつつ、強度の高い負例から明確な嗜好境界を学習することが可能となる。また、BPR 損失の計算には動的に修正されたグラフではなく、元の観測されたインタラクション \mathcal{E}^+ および \mathcal{E}^- を利用する点に留意されたい。これは、最適化プロセス中にラベルを動的に反転させると損失関数に不連続性が生じ、学習が不安定化するためである。

3.3.2 修正済みグラフ対照学習

推薦システムにおける標準的な対照学習では、未観測であるが潜在的に好まれる可能性のあるアイテムを負例として扱い、それらの表現を不当に遠ざけてしまうことが多い。この問題に対処するため、本手法では動的に修正されたグラフ \mathbf{A}' を活用し、ユーザの嗜好と明示的な嫌悪を対比させるアイテム側の対照学習タスクを導入する。具体的には、個々のユーザのインタラクション履歴に基づいてアイテムペアを構築する。ユーザ u に対する修正済みの正および負のアイテム集合をそれぞれ $\mathcal{P}_u = \{i \mid A'_{ui} = 1\}$ および $\mathcal{N}_u = \{k \mid A'_{uk} = -1\}$ とする。各ユーザ u について、 \mathcal{P}_u からアンカーアイテム i と正のアイテム j ($i \neq j$) をサンプリングし、同一ユーザが正のインタラクションを行ったアイテム間に引力を働かせる。逆に、負のアイテム k を部分集合 $\mathcal{S}_u^- \subset \mathcal{N}_u$ からサンプリングし、好まれる

表 1: データセットの統計情報

Dataset	#Users	#Items	#Interactions	Ratio (Pos:Neg)
Amazon-CDs	51,267	46,464	895,266	1 : 0.22
Amazon-Music	3,472	2,498	49,875	1 : 0.25
KuaiRand	16,974	4,373	263,100	1 : 1.25

アイテムと明示的に嫌悪されるアイテムとの間に反発力を働かせる。

前節で定義した層平均アイテム表現 ($\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k$) を用いて、修正済みグラフにおける対照損失は以下の式で定式化する。

$$\mathcal{L}_{CL} = - \sum_{u \in \mathcal{U}} \sum_{\substack{i,j \in \mathcal{P}_u \\ i \neq j}} \log \frac{\exp(s(\mathbf{e}_i, \mathbf{e}_j)/\tau)}{\exp(s(\mathbf{e}_i, \mathbf{e}_j)/\tau) + \sum_{k \in \mathcal{S}_u^-} \exp(s(\mathbf{e}_i, \mathbf{e}_k)/\tau)} \quad (12)$$

ここで、 $s(\cdot, \cdot)$ はコサイン類似度を示し、 τ は温度パラメータである。重要な点として、このサンプリングが修正済みグラフ \mathbf{A}' 上で行われることである。これにより、潜在的な偽陰性が \mathcal{P}_u に再分類されてユーザの好むアイテムへと正しく引き寄せられる一方で、真の嫌悪のみが \mathcal{N}_u に保持され信頼性の高い反発シグナルを提供する。この目的関数により、ユーザの嗜好と嫌悪の両方を正確に反映した、識別力の高いアイテム埋め込み空間が構築される。

最終的な目的関数は、これらの損失を組み合わせたものとして以下のように定式化する。

$$\mathcal{L} = \mathcal{L}_{bpr} + \lambda_{CL} \mathcal{L}_{CL} + \lambda_{reg} \|\Theta\|^2, \quad (13)$$

ここで、 Θ はすべての学習可能パラメータを表し、 λ_{CL} および λ_{reg} はそれぞれ対照学習および L_2 正則化の重みを制御するハイパーパラメータである。

4 評価実験

本節では、3つの実世界データセットを用いて SANE の有効性を検証する。具体的には、(1) 既存手法に対する全体的な性能比較、(2) 各モジュールの性能への寄与度を検証するアブレーションスタディ、(3) 主要ハイパーパラメータの感度分析、(4) 学習過程における正負の識別能力の分析、(5) 学習された埋め込み空間の可視化の観点から評価を行う。

4.1 実験設定

4.1.1 データセット

実験には、Amazon-CDs、Amazon-Music [22]、および KuaiRand [10] の3つの公開データセットを使用した。各データセットの統計量を表1に示す。Amazon-CDs および Amazon-Music はユーザレビューデータであり、先行研究に従い、評価値4以上を正のフィードバック、4未満を負のフィードバックとして扱った。KuaiRand はショート動画プラットフォームにおけるユーザ行動ログであり、クリック行動を正のインタラクション、動画の無視などの露出済み行動を負のインタラクションとして分類した。すべてのデータセットに対して5-core フィ

表 2: ベースラインとの性能比較. 最良結果を太字で示す. また, OOM はメモリ不足, *は $p < 0.05$ での統計的有意性を示す.

	Model	Amazon-CDs		Amazon-Music		KuaiRand	
		Recall@20	NDCG@20	Recall@20	NDCG@20	Recall@20	NDCG@20
Unsigned Graph-based RS	LightGCN [11]	0.1317	0.0772	0.2734	0.1608	0.1195	0.0579
	SGL [31]	0.1355	0.0804	0.2919	0.1743	0.1127	0.0561
	NCL [19]	0.1269	0.0742	0.2935	0.1749	0.1326	0.0651
	SimGCL [35]	0.1297	0.0779	0.2402	0.1479	0.0868	0.0440
	XSimGCL [34]	0.1311	0.0783	0.2724	0.1637	0.1078	0.0540
	LightGCL [1]	0.1264	0.0731	0.2149	0.1256	0.0881	0.0431
Sign-aware Graph-based RS	SiReN [26]	0.1364	0.0798	0.2936	0.1755	0.1153	0.0570
	Pone-GNN [21]	OOM	OOM	0.2751	0.1657	OOM	OOM
	SIGformer [5]	0.1420	0.0840	0.3030	0.1812	0.1489	0.0720
	SANE (Ours)	0.1490*	0.0879*	0.3222*	0.1880*	0.1563*	0.0755*
Improv.	+4.92%	+4.64%	+6.34%	+3.75%	+4.97%	+4.86%	

ルタリングを適用し, 訓練・検証・テストデータを 7:1:2 の比率でランダムに分割した.

4.1.2 評価指標

推薦性能の評価には, 推薦システムにおける標準的な Top- K 指標である Recall@ K および NDCG@ K を採用した. Recall@ K は Top- K リスト中の関連アイテムの割合を測定し, NDCG@ K は上位に配置されたアイテムにより高い重みを付与することでランキング品質を評価する. また, 先行研究 [11, 34, 35] に従い $K = 20$ とした. すべての実験結果は 5 回の独立した実行の平均値として報告する.

4.1.3 ベースライン

SANE の有効性を包括的に検証するため, 以下の 2 カテゴリの最先端ベースラインと比較した.

1) **符号なしグラフベース推薦手法**: 正のインタラクションのみを利用する代表的な手法を採用した.

- **LightGCN** [11]: 特徴変換と非線形性を排除し, 近傍集約に特化した古典的 GNN.
- **SGL** [31]: グラフ構造を拡張して補助的な対照ビューを生成する自己教師あり学習手法.
- **NCL** [19]: 埋め込みクラスタリングから導出される意味的近傍を活用する対照学習手法.
- **SimGCL** [35]: ノード埋め込みにランダムノイズ摂動を加えてデータ拡張を行う対照学習手法.
- **XSimGCL** [34]: SimGCL を拡張し, 層ごとの摂動による効率的な学習を実現するフレームワーク.
- **LightGCL** [1]: 特異値分解を用いたグラフ拡張による効率的な対照学習モデル.

2) **符号認識型グラフベース推薦手法**: 正負両方のインタラクションを利用する手法を採用した.

- **SiReN** [26]: 正グラフと負グラフから個別に得た埋め込みを注意機構で統合する符号認識型手法.
- **Pone-GNN** [21]: 正負のグラフ上で個別にメッセージパッ

シングを行い, 対照学習で統合する GNN.

- **SIGformer** [5]: 符号付きスペクトルエンコーディングとパスエンコーディングを活用する Transformer ベースのモデル.

4.1.4 パラメータ設定

SANE の最適化には Adam [16] を採用し, グリッドサーチによりハイパーパラメータを探索した. 埋め込み次元は先行研究に合わせて $d = 64$ とした. 学習率を $1e-2$, 重み減衰を $1e-4$, 対照学習の温度パラメータを $\tau = 5e-2$, Transformer 層数を $L = 3$, バッチサイズを 1024 に設定した. 偽陽性フィルタリング率 ρ_{pos} , 偽陰性修正率 ρ_{neg} , および対照学習の重み λ_{CL} は $\{0, 1e-3, 5e-3, 1e-2, 5e-2, 1e-1\}$ の範囲で探索した. 学習中は 20 エポックごとに検証セットで評価を行い, 検証性能が 10 回連続で改善しない場合に学習を打ち切る早期終了戦略を適用した.

4.2 全体的な性能比較

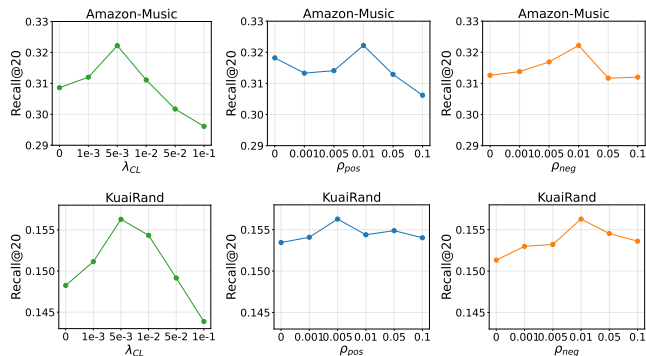
表 2 に全手法の性能比較結果を示す. SANE はすべてのデータセットにおいて全ベースラインを有意に上回り, Amazon-Music では Recall@20 で最大 6.34% の改善を達成した. これらの結果から, 以下の知見が得られた.

符号なし対照学習手法に対する優位性: SANE は XSimGCL や LightGCL などの最先端の符号なしグラフ手法を一貫して上回った. この結果は, 正のグラフを拡張するだけでは識別力のある埋め込み空間の構築には不十分であり, ポジティブな嗜好と構造的に裏付けられた明示的な嫌悪とを対比させることが不可欠であることを示している.

既存の符号認識型手法に対する有効性: 既存の符号認識型手法との比較により, 負のフィードバックの信頼性の違いを考慮することの重要性が確認された. 従来の符号付きモデルはすべての負のシグナルを一律に絶対的な拒絶として扱うため埋め込み表現の歪みが生じるのに対し, SANE は信頼度を動的に評価し不確実な負のエッジを重み付けすることで, 真のユーザ嗜好空

表 3: アブレーションスタディの結果

Method	Amazon-Music		KuaiRand	
	Recall@20	NDCG@20	Recall@20	NDCG@20
Baseline (SIGformer)	0.3030	0.1812	0.1489	0.0720
w/o CL	0.3095	0.1807	0.1501	0.0719
w/o Reweighting	0.3146	0.1838	0.1529	0.0742
w/o Neg Estimation	0.3152	0.1870	0.1519	0.0737
w/o Neg Rectification	0.3126	0.1868	0.1513	0.0735
w/o Pos Filtering	0.3183	0.1864	0.1534	0.0744
SANE (Ours)	0.3222	0.1880	0.1562	0.0755

図 2: λ_{CL} , ρ_{pos} , ρ_{neg} に対するハイパーパラメータ感度分析 (上段: Amazon-Music, 下段: KuaiRand)

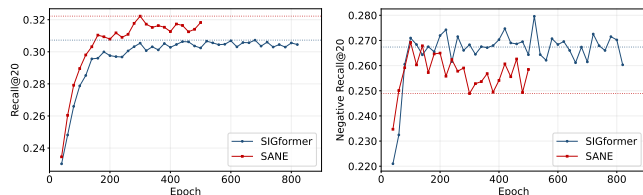
間を適応的に保持している。

深刻なデータ不均衡に対するロバスト性: 正負のインタラクション比率が約 1:1.25 に達する KuaiRand データセットにおいても, SANE は安定した性能向上を達成した. これは, 信頼性の低い負のエッジからの学習リスクが増大する状況下でも, 修正済みグラフ対照学習が曖昧な拒絶に影響されることなくロバストな識別境界を確立できることを実証している.

4.3 アブレーションスタディ

SANE の各モジュールの貢献を検証するため, Amazon-Music および KuaiRand データセット上でアブレーションスタディを実施した. 結果を表 3 に示す.

結果から以下の知見が得られた. まず, ネガティブ信頼度推定モジュール全体を除去した場合 (w/o Neg Estimation) に性能が低下し, 構造的なグラフ洗練が不可欠であることが示された. このプロセス内では, 偽陰性の修正の無効化 (w/o Neg Rectification) および偽陽性のフィルタリングの無効化 (w/o Pos Filtering) のいずれも顕著な性能低下を引き起こした. この結果は, 双方向のグラフ精製が信頼性の高い探索空間を効果的に拡大することを裏付けている. 次に, グラフ修正のみを行い信頼度に基づく BPR 重み付けを省略した場合 (w/o Reweighting) には限定的な改善にとどまった. これは, 単にグラフを書き換えるだけでは不十分であり, 推定された信頼度に基づいて学習シグナルを調整する必要があることを示している. 最後に, 修正済みグラフ対照学習の除去 (w/o CL) も性能を低下させ, 埋め込み表現の識別力向上における本モジュールの役割が検証された. 以上より, すべてのモジュールを統合



(a) Recall@20 (b) Negative Recall@20

図 3: Recall@20 および Negative Recall@20 の学習曲線

することで SANE は精度と網羅性を同時に最大化できることが確認された.

4.4 ハイパーパラメータ感度分析

SANE のロバスト性を評価するため, Amazon-Music および KuaiRand データセット上で主要ハイパーパラメータに対する感度分析を行った. 結果を図 2 に示す.

対照学習の重み λ_{CL} : λ_{CL} の増加に伴い, SANE の性能は初期の改善の後に低下する傾向を示し, いずれのデータセットでも 0.005 で最良となった. 適度な λ_{CL} は埋め込み表現の識別力を高める効果的な構造正則化として機能する一方, 過度に大きな値は BPR 損失の主目的を覆い隠し, 協調フィルタリングプロセスを希薄化させ精度の低下を招く.

グラフ修正率 ρ_{pos} および ρ_{neg} : これらのハイパーパラメータは, 動的にフィルタリングされる偽陽性および修正される偽陰性の割合を制御する. いずれの比率も増加に伴い性能が初期改善の後に低下し, ノイズ軽減と情報保持のバランスにおける役割を反映している. 最適値は ρ_{pos} が Amazon-Music で 0.01, KuaiRand で 0.005 であったのに対し, ρ_{neg} はいずれのデータセットでも 0.01 で一貫していた. 誤クリックや文脈の不一致による修正が必要なインタラクションはごく一部であり, これらの比率を過度に増加させると真の嗜好を誤って除去し, 真の嫌悪を偽陽性へと変換してしまう. これらの結果は, データセットの特性に応じて不確実なフィードバックの影響を柔軟に調整する SANE の適応性を裏付けている.

4.5 学習過程の分析

Amazon-Music データセット上で, Recall@20 および独自指標である Negative Recall@20 の学習曲線を可視化した. Negative Recall@20 は, ユーザが明示的に低評価したアイテムのうち誤って推薦されたアイテムの割合を表し, 以下のように定義される.

$$\text{Negative Recall} = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{|\hat{\mathcal{I}}_u \cap \mathcal{E}_u^-|}{|\mathcal{E}_u^-|} \quad (14)$$

ここで, $\hat{\mathcal{I}}_u$ はユーザ u に対して生成された Top-20 推薦リスト, \mathcal{E}_u^- はユーザ u が明示的に低評価したアイテムの集合を表す.

図 3 に結果を示す. SIGformer は学習中に正のアイテムに対する推薦性能を向上させる一方で, 回避すべき負のアイテムの推薦も同時に増加させている. これは, SIGformer がすべての負のアイテムを一律に扱い, 曖昧な負のシグナルに誤導されるためである. 対照的に, SANE は正のアイテムの推薦性能を向

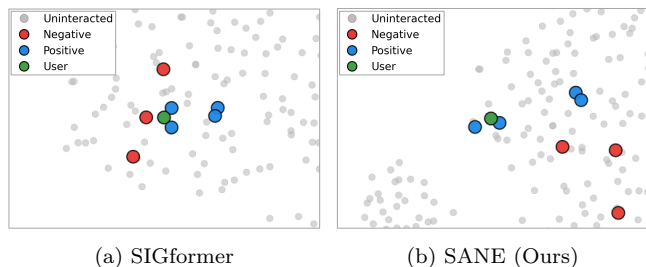


図 4: 学習済み埋め込み空間の t-SNE 可視化

上させつつ、負のアイテムの推薦を抑制することに成功している。この結果は、信頼度に基づく重み付けと動的グラフ修正が曖昧な負のシグナルの影響を効果的に抑制し、真の嫌悪を正しく遠ざけていることの証拠である。明示的に嫌悪されたアイテムを誤って推薦することはユーザー体験を著しく損なうため、提案手法は実用的な展開に適した有効なシステムであるといえる。

4.6 ケーススタディ

SANE が表現学習に与える影響を直感的に理解するため、ランダムに選択したユーザーとそのインタラクション済み・未インタラクションアイテムの学習済み埋め込み空間を t-SNE により可視化した。図 4 に示すように、SIGformer では正のアイテムと負のアイテムが混在し、明示的に嫌悪されたアイテムがユーザー埋め込みの近傍に配置されている。これは、信頼性の異なる負のフィードバックを一律に絶対的な拒絶として扱うことで生じる埋め込み表現の歪みに起因し、信頼性の低い負のフィードバックに最適化が誤導され、強い嫌悪を示すアイテムが不適切にユーザー近傍に留まる結果となっている。対照的に、SANE は明確な空間的分離を持つ識別力の高い埋め込み空間を構築しており、正のアイテムはユーザー周辺に密集する一方、負のアイテムは周辺部へと反発されている。この可視化結果は、インタラクションの信頼度に基づいてグラフ構造を適応的に修正し、好まれるアイテムと真の嫌悪との間に反発力を働かせることで真のユーザー意図を反映した嗜好空間を構築する、提案手法の構造認識型最適化の有効性を裏付けている。

4.7 計算量の分析

SANE の学習コストは 3 つの要素から構成される。符号付きグラフエンコーダが 1 エポックあたり $O((M+N)d\hat{N})$ 、構造認識型ネガティブ信頼度推定が $O((|\mathcal{E}^+| + |\mathcal{E}^-|)d + |\mathcal{E}| \log |\mathcal{E}|)$ 、信頼度に基づく BPR 損失と対照学習損失からなる構造認識型最適化が $O(Bd + |S|d)$ である。ここで、 \hat{N} はサンプリングされる近傍ノードの平均数、 B はバッチサイズ、 $|S|$ はサンプリングされる対照ペア数を表す。信頼度推定モジュールは T エポックごとの周期的な間隔でのみ実行されるため、1 エポックあたりの償却コストは T 分の 1 に削減される。また、すべての追加計算は学習フェーズに限定され、推論時には SANE は標準的な内積スコアリング関数に帰着するため、追加的な遅延は発生しない。

5 関連研究

5.1 符号認識型推薦

負のフィードバックの統合は、現代の推薦システムにおいて重要な課題となっている。初期の符号認識型手法 [7, 13, 14, 18, 21, 26] は、符号付き二部グラフを正と負のサブグラフに分割し、それぞれ個別のエンコーダで処理する。一方、近年の Transformer ベースの手法 [4, 5] は、符号付きスペクトルエンコーディングを通じて大域的な構造情報を捉え、最先端の性能を達成している。しかし、既存の符号認識型手法は観測された負のインタラクションを一律に真の負例として扱い、負のフィードバックの信頼性の違いを見落としている。暗黙的フィードバックに対するデノイズング手法 [29, 32] がこの問題に対処し得る可能性はあるが、それらは未観測インタラクション中の疑似負例を対象としており、明示的に観測された負のインタラクションの信頼性を適応的に扱う既存手法は存在しない。

5.2 推薦システムにおける対照学習

データの疎性を軽減するため、対照学習は推薦システムにおいて広く採用されている。SGL [31] はグラフ構造の拡張により補助的なビューを生成し、SimGCL [35] および XSimGCL [34] は埋め込み空間に直接ノイズ摂動を適用する。LightGCL [1] は特異値分解を用いた構造的対比を行う。符号付き推薦においては、Pone-GNN [21] が正負両方のフィードバックに対照学習を適用している。しかし、これらの標準的な対照学習手法は未観測アイテムを負例としてサンプリングするのが一般的であり、潜在的に好まれるアイテムの表現を不当に遠ざけ、学習されたユーザー嗜好空間を歪めるという問題がある。

6 結論

本研究では、負のフィードバックの信頼性の違いに着目した符号付きグラフ推薦フレームワーク SANE を提案した。既存モデルがすべての負のインタラクションを一律に扱うのとは異なり、SANE は近傍構造から導出される構造的類似度に基づいて負のエッジの信頼性を動的に定量化する推定機構を導入した。潜在的な偽陰性の影響を適応的に軽減し信頼度に基づく重み付け学習を適用することで、信頼性の異なるインタラクションからのロバストな表現学習を可能にした。さらに、構造認識型のアイテム側対照学習を統合し、高信頼度の構造情報を活用して埋め込み空間における正負の嗜好の識別を実現した。3 つの実世界データセットを用いた包括的な実験により、SANE が Recall@20 で最大 6.34% の改善を達成し、提案手法の有効性を実証した。

謝辞

本研究の一部は、ダイキン工業株式会社の助成を受けたものである。ここに記して謝意を表す。

文 献

- [1] Xuheng Cai, Chao Huang, Lianghao Xia, and Xubin Ren. Lightgcl: Simple yet effective graph contrastive learning for recommendation. *arXiv preprint arXiv:2302.08191*, 2023.
- [2] Jiawei Chen, Hande Dong, Yang Qiu, Xiangnan He, Xin Xin, Liang Chen, Guli Lin, and Keping Yang. Autodebias: Learning to debias for recommendation. In *SIGIR*, pp. 21–30, 2021.
- [3] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems*, Vol. 41, No. 3, pp. 1–39, 2023.
- [4] Jinsong Chen, Gaichao Li, John E Hopcroft, and Kun He. Signgt: Signed attention-based graph transformer for graph representation learning. *arXiv preprint arXiv:2310.11025*, 2023.
- [5] Sirui Chen, Jiawei Chen, Sheng Zhou, Bohao Wang, Shen Han, Chanfei Su, Yuqing Yuan, and Can Wang. Sigformer: Sign-aware graph transformer for recommendation. In *SIGIR*, pp. 1274–1284, 2024.
- [6] Fan RK Chung. *Spectral graph theory*, Vol. 92. American Mathematical Soc., 1997.
- [7] Tyler Derr, Yao Ma, and Jiliang Tang. Signed graph convolutional networks. In *ICDM*, pp. 929–934. IEEE, 2018.
- [8] Jingtao Ding, Yuhan Quan, Xiangnan He, Yong Li, and Depeng Jin. Reinforced negative sampling for recommendation with exposure data. In *IJCAI*, pp. 2230–2236. Macao, 2019.
- [9] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. Graph neural networks for social recommendation. In *WWW*, pp. 417–426, 2019.
- [10] Chongming Gao, Shijun Li, Yuan Zhang, Jiawei Chen, Biao Li, Wenqiang Lei, Peng Jiang, and Xiangnan He. Kuairand: An unbiased sequential recommendation dataset with randomly exposed videos. In *CIKM*, pp. 3953–3957, 2022.
- [11] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *SIGIR*, pp. 639–648, 2020.
- [12] Fritz Heider. Attitudes and cognitive organization. *The Journal of psychology*, Vol. 21, No. 1, pp. 107–112, 1946.
- [13] Junjie Huang, Huawei Shen, Qi Cao, Shuchang Tao, and Xueqi Cheng. Signed bipartite graph neural networks. In *CIKM*, pp. 740–749, 2021.
- [14] Junjie Huang, Ruobing Xie, Qi Cao, Huawei Shen, Shao-liang Zhang, Feng Xia, and Xueqi Cheng. Negative can be positive: Signed graph neural networks for recommendation. *Information Processing & Management*, Vol. 60, No. 4, p. 103403, 2023.
- [15] Olivier Jeunen. Revisiting offline evaluation for implicit-feedback recommender systems. In *RecSys*, pp. 596–600, 2019.
- [16] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [17] Chaoliu Li, Lianghao Xia, Xubin Ren, Yaowen Ye, Yong Xu, and Chao Huang. Graph transformer for recommendation. In *SIGIR*, pp. 1680–1689, 2023.
- [18] Yu Li, Meng Qu, Jian Tang, and Yi Chang. Signed laplacian graph neural networks. In *AAAI*, Vol. 37, pp. 4444–4452, 2023.
- [19] Zihan Lin, Changxin Tian, Yupeng Hou, and Wayne Xin Zhao. Improving graph collaborative filtering with neighborhood-enriched contrastive learning. In *WWW*, pp. 2320–2329, 2022.
- [20] Ziyang Liu, Chaokun Wang, Jingcao Xu, Cheng Wu, Kai Zheng, Yang Song, Na Mou, and Kun Gai. Pane-gnn: Unifying positive and negative edges in graph neural networks for recommendation. *arXiv preprint arXiv:2306.04095*, 2023.
- [21] Ziyang Liu, Chaokun Wang, Shuwen Zheng, Cheng Wu, Kai Zheng, Yang Song, and Na Mou. Pone-gnn: integrating positive and negative feedback in graph neural networks for recommender systems. *ACM Transactions on Recommender Systems*, Vol. 3, No. 2, pp. 1–23, 2025.
- [22] Julian John McAuley and Jure Leskovec. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. In *WWW*, pp. 897–908, 2013.
- [23] Yunzhu Pan, Chen Gao, Jianxin Chang, Yanan Niu, Yang Song, Kun Gai, Depeng Jin, and Yong Li. Understanding and modeling passive-negative feedback for short-video sequential recommendation. In *RecSys*, pp. 540–550, 2023.
- [24] Minju Park and Kyogu Lee. Exploiting negative preference in content-based music recommendation with contrastive learning. In *RecSys*, pp. 229–236, 2022.
- [25] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618*, 2012.
- [26] Changwon Seo, Kyeong-Joong Jeong, Sungsu Lim, and Won-Yong Shin. Siren: Sign-aware recommendation using graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 35, No. 4, pp. 4729–4743, 2022.
- [27] Jiliang Tang, Huiji Gao, Huan Liu, and Atish Das Sarma. etrust: Understanding trust evolution in an online world. In *SIGKDD*, pp. 253–261, 2012.
- [28] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, Vol. 30, , 2017.
- [29] Wenjie Wang, Fuli Feng, Xiangnan He, Liqiang Nie, and Tat-Seng Chua. Denoising implicit feedback for recommendation. In *WSDM*, pp. 373–381, 2021.
- [30] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. Neural graph collaborative filtering. In *SIGIR*, pp. 165–174, 2019.
- [31] Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. Self-supervised graph learning for recommendation. In *SIGIR*, pp. 726–735, 2021.
- [32] Haibo Ye, Lijun Zhang, Yuan Yao, and Sheng-Jun Huang. Denoised graph collaborative filtering via neighborhood similarity and dynamic thresholding. *IEEE Transactions on Big Data*, 2024.
- [33] Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? *Advances in neural information processing systems*, Vol. 34, pp. 28877–28888, 2021.
- [34] Junliang Yu, Xin Xia, Tong Chen, Lizhen Cui, Nguyen Quoc Viet Hung, and Hongzhi Yin. Xsingcl: Towards extremely simple graph contrastive learning for recommendation. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 36, No. 2, pp. 913–926, 2023.
- [35] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. Are graph augmentations necessary? simple graph contrastive learning for recommendation. In *SIGIR*, pp. 1294–1303, 2022.

グループ単位の関連度評価に基づく 知識グラフ探索アルゴリズムの設計と評価

八尾 正剛[†] 福井 健太[†] LiGuangcan[†] 鬼塚 真[†]

[†] 大阪大学 〒 565-0871 大阪府吹田市山田丘 1-1

E-mail: †{yao.seigo,fukui.kenta,guangcan.li,onizuka}@ist.osaka-u.ac.jp

あらまし LLMの推論精度を向上する技術として、GraphRAGにおける知識グラフ上での情報検索の重要性が高まっている。既存研究では、与えられた質問を構成する各質問エンティティをクエリとして検索して得られるノード集合をグループとして表現し、各グループから少なくとも1つのノードを含み、かつノード間の接続コストの総和が最小となるサブグラフを決定する問題として Group Steiner Tree (GST) に基づく手法が広く用いられている。しかし、GSTの目的関数は接続コストの最小化に限定されているため、質問エンティティに対するノードの適合度という本質的な情報を活用したサブグラフ選択を行うことができない。そこで本研究では、各グループに属するノードの適合度を単調劣モジュラ関数で集約した値を目的関数に組み込むことで、本問題を定式化する。さらに、総当たりでは実行時間が長くなるため、事前計算と複数の評価指標に基づく段階的な局所探索を導入し、高速に近似解を探索する。実験では、3つのデータセットと複数の設定における実験において、既存手法よりも高い精度を達成した。

キーワード 知識グラフ, 局所探索法

1 序 論

近年、知識グラフを活用し、グラフ上の関係をもとに情報を検索・統合して検索結果や質問応答の形で関連情報を提示する知識集約型タスクが注目されている [1, 2]。このような知識グラフ活用の枠組みは、知識グラフを用いた情報検索 [3, 4] や推薦 [5] に加え、医療・バイオ分野の質問応答 [6]、サイバーセキュリティにおける情報分析・質問応答 [7]、法務領域の文書QA [8]、製造業の故障モード影響解析 [9] など様々な分野で応用されている。

しかし、知識集約型タスクで用いられる知識グラフは一般に大規模であり、入力クエリに関連する情報はグラフ全体に散在するため、クエリに適合する根拠サブグラフをどのように選択するかが、検索精度と計算効率の両面から重要な課題となっている [3, 10, 11]。

実運用のクエリには、対象となるエンティティを指す語句に加えて、エンティティ間に求める関係や条件を表す語句が併存し、複数のエンティティ・関係・制約を同時に含むことが多い [12]。そのため、各クエリ要素に対応する候補ノードのクエリ要素への適合度を考慮しつつ、要素ごとの取りこぼしが生じないように各要素を被覆する必要がある [11, 13]。このようなトレードオフを定量的に扱うため、根拠サブグラフ抽出はグラフ上の最適化問題として定式化される [14-16]。

根拠サブグラフ抽出の代表的な定式化として、質問を構成する質問エンティティごとに対応する候補ノード集合をグループとして定義し、すべてのグループを被覆しつつ接続コストが最小となる部分グラフを求める Group Steiner Tree (GST) が知られている [13, 16]。GSTはクエリを構成する要素を必ず含むという制約を表現できる一方で、目的関数が接続コストの最

小化に限定されるため、候補ノードの質問エンティティへの適合度を目的関数に直接取り込めないという課題がある。また、ノードに質問との適合度を表す重みを prize として与え、接続コストとのトレードオフを最適化することが可能な定式化として Prize-Collecting Steiner Tree (PCST) [17, 18] が知られている。しかし、PCSTではグループ被覆を制約に含んでいないため、クエリにおける要素が欠落してしまう場合がある。更に、ノードの重みと接続コストの両方を最適化する新たな定式化の方法として Node-Weighted Group Steiner Tree (NW-GST) が知られている [19]。しかし、NW-GSTではノード重みを線形加算でしか評価することができないため、同一グループ内でノードを追加で選ぶことによる利得は選択済みのノードに依存せず常に一定となり、情報検索分野で知られている限界減性質 (既に選択した情報と類似した情報を追加しても得られる価値が小さくなる性質) [20] を表現できないという課題がある。

そこで本研究では、そのような限界減性質を単調劣モジュラ関数として目的関数に組み込むことで、本問題を定式化する。具体的には、各グループに対して解に含まれるノードの適合度を、単調劣モジュラ関数で集約し、これを接続コストと同時に最適化することで、グループ被覆を満たしつつ、各グループ内で高適合度ノードを優先して選択する根拠サブグラフ抽出を実現する。また、本問題を単純な総当たりで厳密に解くことは計算量の観点から現実的ではないため、本研究では目的関数を近似した複数の評価指標に基づく段階的な局所探索を導入し、探索空間を絞り込みながら高速に近似解を求める手法を提案する。

本稿の貢献を以下にまとめる。

- 知識集約型タスクにおける根拠サブグラフ抽出に対し、グループ被覆とグループ内適合度を単調劣モジュラ関数で集約した値を、同時に扱う最適化問題を定式化した。
- 総当たり探索の計算量的困難さに対し、複数の評価指標を

用いた段階的な局所探索により、高速に近似解を探索するアルゴリズムを提案した。

- 複数データセット・複数設定において実験を行い、精度と実行時間の両立という観点から提案手法の有効性を検証した。

2 関連研究

本研究が扱う根拠サブグラフ抽出は、関連度の高いノードの選択とそれらの構造的接続コストを同時に最適化する Steiner Tree 系の組合せ最適化問題の拡張として位置付けられる。このタイプの問題は、知識グラフ探索、キーワード検索、QA などで広く用いられている。代表的な定式化として、グループ被覆制約を持つ Group Steiner Tree (GST), ノード適合度とエッジコストのトレードオフを扱う Prize-Collecting Steiner Tree (PCST), およびノード重みを考慮する Node-Weighted Group Steiner Tree (NW-GST) が知られている。以下では、これらの定式化を整理し、特性と限界を述べる。

Group Steiner Tree (GST) : GST は、グラフ上でグループとして与えられる複数の候補集合をそれぞれ少なくとも 1 点ずつ含み、エッジ重みの和が最小となる木を求める問題である。クエリ検索では、クエリ語ごとに類似したノード集合をグループとみなし、それらを結ぶ最小コストの木を返す定式化が古典的に用いられてきた [10, 14, 21].

知識グラフ探索の分野でも、クエリに関連するエンティティを結ぶ根拠サブグラフを、GST や topk-GST から求める手法が用いられている [10, 22]. また、QA では、複数の文書から抽出した関係をノード・エッジとして構成した擬似的な知識グラフから、GST を用いてクエリに対応した複数の根拠を接続し、回答生成するシステムが提案されている [13, 23].

Prize-Collecting Steiner Tree (PCST) : PCST は、各ノードにクエリに対する適合度に応じた prize, 各エッジにコストが与えられたときに、含まれるノードの適合度の和とエッジコストのトレードオフが最適化されるような木を求める問題である。Graph RAG では、大規模グラフから LLM に投入可能なサイズまで情報を圧縮しつつ、適合度の高いノードをなるべく多く含み、かつ構造的にまとまったサブグラフを抽出する目的で PCST 定式化を採用したものが知られている [11]. また、PCST は、RAG 用途のサブグラフ抽出でも利用されている。例えば、対話システム向けに、文脈関連度を報酬として与え、最小サイズで最大関連のサブグラフを得るために PCST として解く手法が提案されている [24]. また、マルチモーダル知識グラフ QA でも、関連性と構造的な一貫性を担保するサブグラフ抽出として PCST が用いられている [25].

Node-Weighted Group Steiner Tree (NW-GST) : NW-GST は、GST のようなグループ被覆制約を保ちつつ、エッジ重みに加えてノード重みもコストとして取り込み、ノードとエッジのコストの総和が最小となる木を求める問題である。このような設定は、ノードそのものにコストが設定されるような課題で有用である [26].

NW-GST のアルゴリズムの研究としては、GST における既存近似手法を拡張したアルゴリズムなどを与える研究が知られている [19]. NW-GST は、リレーショナル DB からの情報検索や、ソーシャルネットワーク、知識グラフからの情報抽出に用いることができる [19].

既存定式化の課題: 既存の定式化である GST, PCST, NW-GST には、それぞれ次のような課題がある。GST はノードの適合度を目的関数に含んでいないため、ノードの適合度も考慮した上でクエリに対して最適な木を選択することができない。PCST はグループ被覆を制約として含んでいないため、各クエリ要素に対する情報の取りこぼしが発生することがある。また、NW-GST はグループごとに選択したノードの適合度を単純な加算で評価するため、情報検索において基本的な考えである限界通減性質を考慮した木を選択することができない。これらの課題を解決するため、4 章で新たな定式化を提案する。

3 事前知識

知識グラフを $G = (V, E)$ と表す。ここで、知識グラフは無向グラフであり、 V はノード集合、 E はエッジ集合である。各エッジ $e \in E$ には非負の接続コスト $c(e) \in \mathbb{R}_{\geq 0}$ を割り当て、サブグラフ $T = (V_T, E_T)$ の接続コストを

$$\text{Cost}(T) = \sum_{e \in E_T} c(e) \quad (1)$$

と定義する。また、ノード $u, v \in V$ 間の最短距離は $d(u, v)$ と表し、ノード v から集合 H への距離を $d(v, H) = \min_{u \in H} d(v, u)$ と書く。 $d(v, H)$ は multi-source dijkstra を用いて求める [27].

a) 質問エンティティに対応するノードグループ

質問 q は、 m 個の質問エンティティから構成されるとする。各質問エンティティ q_i ($1 \leq i \leq m$) に対して、類似度関数 $\text{sim}(\cdot, \cdot)$ に基づき、対応する候補ノード集合をノードグループ G_i として

$$G_i = \text{TopK}(V, v \mapsto \text{sim}(q_i, v), k)$$

と定義する。このとき、質問エンティティ q_i とノード v の間で算出される適合度 $\text{sim}(q_i, v)$ を、 $q[i]$ に関するノード v の prize と呼び、一般的にサブクエリとノードの埋め込み類似度や BM25 のスコア等から導出される。

b) グループ被覆条件

質問を構成する全ての質問エンティティを網羅した根拠サブグラフ T を選択するために、以下で定義されるグループ被覆を条件として導入する。

$$V_T \cap G_i \neq \emptyset \quad (\forall i \in \{1, \dots, m\}) \quad (2)$$

この条件は各グループから少なくとも 1 ノードを含むことに相当し、質問を構成する全ての質問エンティティに対応する根拠の欠落を抑える目的で導入される。

このように、入力となる知識グラフ $G = (V, E)$, グループ G_i , グループ被覆条件を定義する。また、グループ G_i に対するノード $v \in V$ の prize を $p_i(v)$ とする。

4 提案手法

本研究では、グループ被覆制約を満たしつつ、各グループ内で適合度の高いノードを優先して選択し、かつノード間の接続コストを同時に考慮する根拠サブグラフ抽出問題を定式化する。但し、これらを同時に最適化する組合せ最適化問題は計算量的に困難であり、厳密解法は大規模グラフでは現実的でない。そこで本研究では、実用的な計算時間内で高品質な解を得ることを目的として、計算効率と解品質の両立を図る。具体的には、制約付き組合せ最適化で広く用いられる局所探索を用いてグループ被覆制約を保ったまま解を逐次改善するアプローチを採用し、局所探索における初期解の構成および評価機構を以下の方針に基づいて具体化する。(1) prize が高く、かつ他のグループに近い有望なノードを各グループごとに選択し、初期解とする。(2) 局所探索では、ノードの適合度と現在の選択集合との距離の両方を考慮した有望度指標に基づいて探索候補を優先順位付けする。(3) 解の評価においては、厳密な Steiner 木コストの計算は高コストであるため、軽量な近似評価からより厳密な評価へと移行する 2 段階の評価指標を用いる。

以降、4.1 節で根拠サブグラフ抽出問題を定式化し、4.2 節で、高速に近似解を得るための局所探索アルゴリズムを述べる。

4.1 提案問題の定式化

本節では、入力として与えられる知識グラフ、質問エンティティに対応するノードグループ、およびノード適合度に基づき、本研究で扱う根拠サブグラフ抽出問題を定式化する。

本研究の目的は、(i) 式 (2) で定義されたグループ被覆制約を満たしつつ、(ii) 各グループに対して高い適合度をもつノードを優先的に含み、(iii) 接続コストが大きくなる根拠サブグラフ $T = (V_T, E_T)$ を抽出することである。

まず、prize に関する条件 (ii) と接続コストに関する条件 (iii) を表現する目的関数を以下のように定義する。

定義 1 (目的関数). 連結サブグラフ $T = (V_T, E_T)$ に対し、目的関数 $F(T)$ を以下のように定義する。

$$F(T) = \text{Cost}(T) - \lambda \sum_{i=1}^m S_i(T) \quad (3)$$

但し、 $\lambda > 0$ は各グループ G_i に対する寄与度 $S_i(T)$ の総和と接続コストのトレードオフを制御する係数である。各グループ i と候補ノード $v \in G_i$ に対して、クエリ q_i との適合度 (prize) を

$$p_i(v) := \text{sim}(q_i, v) \quad (4)$$

と定義する。さらに、本研究では単調劣モジュラ関数 f を導入し、 $S_i(T)$ を以下で定義する：

$$S_i(T) = f(\{\sum_{v \in V_T \cap G_i} p_i(v)\}). \quad (5)$$

式 (5) は、同一グループ内で複数ノードを追加したときの寄与が限界通減するよう、単調劣モジュラ性に基づいて評価を集約する点に特徴がある。具体的には、任意の集合 $A \subseteq B$ と要

素 $x \notin B$ に対して

$$S_j(A \cup \{x\}) - S_j(A) \geq S_j(B \cup \{x\}) - S_j(B) \quad (6)$$

が成り立つため、既に多くのノードが選ばれているほど追加ノードの増分は小さくなる。この性質により、不要なノードが T に混入することを抑制しつつ、各グループの情報を段階的に補うことが可能である。

次に、グループ被覆条件 (i) を満たすことを制約として、上記の目的関数の最適化問題を以下のように定式化する。

定義 2 (サブグラフ抽出の最適化). $G = (V, E)$ を与えられたグラフとする。本研究では、 G の連結サブグラフ $T = (V_T, E_T)$ に対し、次の最適化問題を考える。

$$T^* = \arg \min_{T \subseteq G} F(T) \quad (7)$$

$$\text{s.t. } V_T \cap G_i \neq \emptyset, \quad \forall i \in \{1, \dots, m\}, \quad (8)$$

T is connected.

4.2 提案アルゴリズム

本節では、定義 2 で示したサブグラフ抽出の最適化問題に対し、高速に近似解を得るための局所探索アルゴリズムを述べる。本アルゴリズムは、目的関数に対してそのノードがどれだけ有望かを数値化して探索候補を決定する指標に基づく初期選択、二段階評価に基づく局所探索、および一括再選択による再初期化、得られたターミナル集合によるサブグラフの構築から構成される。以降では、各グループ G_i から選択されるターミナルノードの集合を $V_T = \{r_1, \dots, r_m\}$ ($r_i \in G_i$) と表わす。

4.2.1 初期解の選択

本アルゴリズムは局所探索を行うため、初期解の選択が最終的に得られる解の品質に大きく影響する。このため定義 1 で示した目的関数に基づいて、各グループごとに prize が高く、かつ他グループとの距離が近い有望なノードを初期解として選択することが重要である。有望なノードを選択する指標として、グループ G_j におけるノード v のスコア

$$B_j(v) = S_j(v)/P_{\max} + \beta A_j(v) \quad (9)$$

を導入し、各グループ G_j ごとに $r_j = \arg \max_{v \in G_j} B_j(v)$ となるノード r_j を初期解として選択することで、初期解 $V_T \leftarrow \bigcup_{j=1}^m \{r_j\}$ を得る。但し、右辺の第一項で使用されている

$$P_{\max} = \max_j \max_{v \in G_j} f_j(\{v\})$$

であり、prize を正規化するために用いる。また、第二項の $A_j(v)$ は他グループへの近さを近似する指標であり、以下のように定義される。

$$A_j(v) = \frac{1}{m-1} \sum_{\ell \neq j} \exp\left(-\frac{D_\ell(v)}{\sigma}\right) \quad (10)$$

σ には $\{D_\ell(v) \mid v \in \bigcup_j G_j, \ell = 1..m\}$ の中央値を用いる。指数関数を用いることで、極端に大きな $D_\ell(v)$ の影響を小さくしつつ、ノード v の他のグループ l への近さを近似する。

4.2.2 二段階評価に基づく局所探索

ここでは、現在のターミナル集合 V_T に対し、以下の3種類の近傍操作を反復する。

- **ADD**(b) : $V_T \leftarrow V_T \cup \{b\}$,
- **REMOVE**(a) : $V_T \leftarrow V_T \setminus \{a\}$,
- **EXCHANGE**(a, b) : **ADD**(b) の後に **REMOVE**(a).

a) 探索候補の絞り込み

各反復ごとに、目的関数の値を改善する可能性の高い有望なノードを探索候補集合 B とする。このような目的関数に対して有用度が高いノード v は、現在のターミナル集合 V_T に加えた場合の木のスコアの増分が大きく、かつ V_T への距離が小さい方が望ましいことから、PIA スコアを次のように設計する。

$$J_j(v, V_T) = \alpha \frac{\Delta S_j(v, V_T)}{P_{\max}} + \beta \exp(-d(v, V_T)/\sigma_T), \quad (11)$$

$$\Delta S_j(v, V_T) = S_j(V_T \cup \{v\}) - S_j(V_T).$$

但し、 $S_j(V_T) = V_T \setminus (V_T \cap G_j)$ であり、 σ_T には $d(\cdot, V_T)$ の中央値を用いる。式 (11) の右辺の第1項はノード v を V_T に加えた場合の木のスコアの増分近似であり、第2項はノード v の V_T に対する近さを表したものであるため、式 (11) の値はノード v の V_T に対する有望度を表す。そして、PIA スコアが各グループで上位 t 以内のノードの集合を探索候補集合 B とする。

b) 代理評価を用いたフィルタリング

式 (1) で示されるサブグラフの接続コストを厳密に計算するには、シュタイナー木の計算が必要となるため、探索候補集合 B が与えられたときに、考えられる近傍操作全てに対して式 (1) 計算すると高コストである。そこで本アルゴリズムでは、安価な代理評価によって改善の見込みが高い操作のみを残す。

ノード $v \in G_j$ を V_T に加える操作 **ADD**(b) については、木コストの増分を

$$\widehat{\Delta C}_{\text{add}}(b) = d(b, V_T)$$

で近似し、寄与度増分は b を含む全グループの限界利得を足し合わせて

$$\widehat{\Delta P}_{\text{add}}(b) = S_j(V_T \cup \{b\}) - S_j(V_T)$$

とする。そして

$$\widehat{\Delta F}_{\text{add}}(b) = \widehat{\Delta C}_{\text{add}}(b) - \widehat{\Delta P}_{\text{add}}(b) \quad (12)$$

の小さい候補 b を少数残す。

REMOVE(a) については、現在のターミナル集合 V_T と $V_T \setminus \{a\}$ の目的関数値を次の $\widehat{F}(V_T)$ で粗く見積もる。

$$\widehat{F}(V_T) = \widehat{C}(V_T) - \sum_{j=1}^m S_j(V_T), \widehat{C}(V_T) = \min_{\ell} \sum_{t \in V_T} D_{\ell}(t) \quad (13)$$

そして \widehat{F} の差分 $\widehat{F}(V_T) - \widehat{F}(V_T \setminus \{a\})$ を現在のターミナル集合 V_T からノード a を取り除いたときのスコアの減少幅とみなし、その値が大きい候補 a を少数残す。EXCHANGE についても同様に、現在のターミナル集合 V_T と $(V_T \setminus \{a\}) \cup \{b\}$ に対する

式 (13) の差分を計算し、その値が大きい候補 (b, a) を少数残す。EXCHANGE は **ADD** と **REMOVE** の組み合わせとして表現できるが、**ADD**, **REMOVE** を個別に行った場合では目的関数値が改善しない場合でも、EXCHANGE として同時に実行すれば目的関数が改善することがあり得るため導入する。

c) 厳密評価と受理

代理評価で残った少数の候補に対してのみ、式 (1) の値を実際に計算し、 F を最も改善する操作を1つだけ受理して V_T を更新する。同様の反復を最大反復回数 R まで、あるいは改善が得られなくなるまで繰り返す。

4.2.3 再初期化

4.2.2 節の局所探索は、目的関数の値が改善する場合のみ解を更新し、解の改悪を許容しない。そのため、探索範囲が狭まり局所解に陥る可能性が高くなっている。この問題を緩和するため、局所探索後に初期解の再選択を行い、その初期解を元に局所探索をもう一度行うことで探索範囲を大幅に変更する。現在のターミナル集合を V_T として、各グループ j に対し1ノードを再選択するスコアを

$$I_j(v) = \alpha \frac{S_j(\{v\})}{P_{\max}} + \beta A_j(v) + \eta A_j^{\text{term}}(v; V_T) \quad (14)$$

と定義する。 $A_j(v)$ は (10) と同様に他グループへの近さを表す項であり、 $A_j^{\text{term}}(v; V_T)$ は現在のターミナル群への近さを表す項であり、

$$A_j^{\text{term}}(v; V_T) = \frac{1}{|S_j(V_T)|} \sum_{t \in S_j(V_T)} \exp(-d(v, t)/\sigma_T) \quad (15)$$

である。これにより、各グループで $I_j(v)$ 最大の候補を選び直して新しい V_T を作り、4.2.2 節の局所探索を再実行し、最適なターミナル集合 V_T^* を探索する。

4.2.4 出力

以上のようにして、得られたターミナル集合 V_T^* をターミナルとして、既存の Steiner tree 近似アルゴリズム [28] を実行することで最終的なサブグラフを出力する。

5 実験

提案手法の優位性を示すため、以下に示す3つの問いを設定して評価実験を実施した。

実験 1: 提案手法は、既存の根拠サブグラフ抽出と比べて、グループ被覆を満たしつつ高い適合度の根拠を選択できるか?

実験 2: 提案アルゴリズムの構成要素はそれぞれ、精度にどの程度寄与するか?

実験 3: ハイパーパラメータは、解の傾向にどのような影響を与えるか?

5.1 実験設定

データセット: 本研究では、Cora, CiteSeer, PubMed の3つの引用ネットワークデータセットを用いる [29]。いずれも論文をノード、引用関係を辺として表現した単一グラフからなり、各ノードには Bag-of-Words に基づく特徴量とトピックラベルが付与されている。これらのノード数、エッジ数、特徴量数は

データセット	ノード数	エッジ数	特徴量数
Cora	2,708	5,429	1,433
CiteSeer	3,327	4,732	3,703
PubMed	19,717	44,338	500

表 1: データセット統計

表 1 のようになっている [30].

グループ構築: m をグループ数, k は各グループが含むノード数とする. グループを構築するために, グラフ上でランダムなノードを m 個選択し, それらをシードノードとする. これらシードノードとの類似度が上位 k 以内であるノードをそれぞれのシードノードを元にしたグループとする.

prize 設定: 各グループ G_j に対してグループ内のノードの, シードノードとの類似度に基づく順位を $r_j(v) \in \{1, \dots, |G_j|\}$ で表す. このとき, グループ G_j におけるノード v の基礎 prize を

$$\tilde{p}_j(v) = \begin{cases} |G_j| - r_j(v) & (v \in G_j), \\ 0 & (v \notin G_j) \end{cases} \quad (16)$$

と定義する. 実験では, この基礎 prize に式 (1) 中でのスケール係数 $\lambda > 0$ を乗じた

$$p(v) = \lambda \tilde{p}(v) \quad (17)$$

を最終的な prize として用いる. ここで, λ は目的関数における prize による利得と接続コストの相対的な重要度を制御するパラメータである. λ を大きくすると, 高 prize ノードを得ることを優先して多少の遠回りを許容する傾向が強まり, 逆に λ を小さくすると, グラフ構造を優先して prize の差を相対的に無視する傾向が強まる. したがって, λ の設定は抽出される根拠サブグラフの性質に大きく影響する.

比較手法: 提案手法の有効性を検証するため, 以下の代表的な手法と比較する.

- Max-Prize: 各グループ G_i から prize が最大のノードを 1 つずつ選択し, それらをターミナルとして Mehlhorn の Steiner 木近似により接続する [28].
- PCST (pcst_fast): PCST の近似実装である pcst_fast を用いる [31]. ただし, PCST は標準形ではグループ被覆制約を持たないため, 各グループを被覆できるようにダミーノードを追加する変換を施し, 変換後の PCST を解くことでグループ制約を実現する.
- NW-GST (exENSteiner): Node-weighted GST の既存アルゴリズムである exENSteiner を用いる [19]. グループごとに劣モジュラ関数で集約された適合度を, サブグラフのスコアとして加算するという提案問題における目的関数を再現するため, ダミーノードを追加する変換をグラフに施した上で解く. 具体的には, まず各グループ G_i で単体での prize が最大のノードを 1 つ選び, その後はすでに選んだ集合 S_i に追加したときの増分

$$f(S_i \cup \{v\}) - f(S_i)$$

が最大となる候補を順に最大 3 個まで選ぶことで, あるノードを選択したときの利得 $\Delta_i(v)$ を段階的に算出する. 残りの候補は, この 3 個の集合に対する増分

$$\Delta_i(v) = f(S_i \cup \{v\}) - f(S_i)$$

により近似する. そして, 各候補 v に対してダミー葉 v^* を追加して v と 0 コストの辺で接続する. v の重みを 0 とし, ダミー葉の重みを

$$w(v^*) = M - \Delta_i(v)$$

と設定することで, $\Delta_i(v)$ が大きい候補ほど重みが小さくなり, 解として選ばれやすくなる.

- NW-GST (FastAPP): NW-GST (exENSteiner) と同様に, ダミーノードを用いたグラフの変換をした上で, NW-GST の FastAPP を用いる [19].
- NW-GST (ImprovAPP): NW-GST (exENSteiner) と同様に, ダミーノードを用いたグラフの変換をした上で, NW-GST の ImprovAPP を用いる [19].

評価指標: 評価指標として目的関数の値および平均順位, 勝率を用いる. 提案問題は最小化問題であるため, 目的関数値が小さいほど良い解である. また, あるアルゴリズムの順位としては, その実験設定において最も目的関数の値が良い解を探索したアルゴリズムの順位を 1 とし, それ以外のアルゴリズムの順位を降順でつける. なお, 目的関数の値が同じであった場合は同じ順位をつける. 勝率は, 複数回の実験において順位が 1 であった割合とする.

ハイパーパラメータ: 実験では, グループサイズ k を 30, グループ数 m を $\{4, 8, 12, 16\}$, 式 (1) 中でのスケール係数である λ を $\{0.5, 1.0, 2.0, 4.0\}$ に変化させ, 各実験設定ごとにシードを 8 通りに変えて実験を行った. 単調劣モジュラ関数としては, 代表的な log 関数, max 関数, sqrt 関数, topk 関数を用いて実験した [32, 33]. 提案手法における局所探索の最大反復回数 R は 20 とした¹.

実験環境: 全ての実験は, Ubuntu 20.04.6 LTS 上のサーバ (HPE ProLiant DL385 Gen10 Plus) で行った. 提案アルゴリズムの実装については Python で行い, NW-GST の実装は C++ で行った, 主要ライブラリとして NumPy, SciPy, NetworkX, scikit-learn, pcst_fast を用いた.

5.2 実験 1

実験 1 では, 提案手法が, 既存の根拠サブグラフ抽出と比べて, グループ被覆を満たしつつ高い適合度の根拠を選択できるかを確認するため, パラメータを $\lambda = 1.0$, $m = 8$, $f = \log$ に固定して実験し, 既存手法と提案手法の比較を行った. 実験結果を表 2 に示す. それぞれの値は 8 回の実験の結果を平均したものである. $\lambda = 1.0$ は, 本研究と同様に prize とコスト両方の最適化をすることで根拠サブグラフを抽出する既存研究を

¹: 早期終了した実験の 95% が 17 ラウンド以内で終了したため $R = 20$ と設定した.

手法	Cora			Citeseer			Pubmed		
	目的関数値 ± 標準偏差	平均順位		目的関数値 ± 標準偏差	平均順位		目的関数値 ± 標準偏差	平均順位	
Max-Prize	0.67 ± 4.48	6.00		9.11 ± 4.07	6.00		4.62 ± 4.13	5.25	
pcst_fast	-6.09 ± 1.77	4.12		-5.54 ± 1.44	3.25		-3.43 ± 1.98	2.38	
exENSteiner	-9.49 ± 2.92	2.69		-5.33 ± 3.36	3.81		-0.61 ± 2.65	3.81	
FastAPP	-5.46 ± 3.05	4.25		-2.36 ± 4.70	4.00		5.74 ± 5.36	5.25	
ImprovAPP	-9.45 ± 2.30	2.81		-6.60 ± 3.14	2.69		-1.27 ± 2.57	3.06	
提案手法	-13.62 ± 2.22	1.12		-8.91 ± 1.89	1.25		-6.37 ± 1.98	1.25	

表 2: 固定したパラメータにおける手法ごとの目的関数値と標準偏差, および平均順位

手法	Cora	Citeseer	Pubmed
Max-Prize	4.02	4.09	4.04
pcst_fast	5.53	5.31	5.22
exENSteiner	3.01	3.11	2.72
FastAPP	4.17	4.13	4.79
ImprovAPP	2.71	2.71	2.64
提案手法	1.56	1.64	1.59

表 3: データセット別の平均順位. 最も高い精度のものを太字にしている.

踏まえて設定した [11]. $m = 8$ は, 本研究での比較手法として用いている NW-GST の既存研究での実験設定を踏まえて設定した [19]. $f = \log$ は, 本研究と同様に, 劣モジュラ関数を用いて効用を表現する既存研究を踏まえて設定した [32]. また, データセットの解の品質に対する影響を確認するため, データセットごとでの既存手法と提案手法の比較を行った. その結果を表 3 に示す. それぞれの値は 512 回の実験の結果を平均したものである. 表 2 から, 提案手法が比較手法に対して高い精度, および安定性を両立していることがわかる. これは, 提案手法が既存手法と異なり, グループごとの prize を集約する劣モジュラ関数を考慮して解を探索できているからだと考えられる. また, 表 3 からすべてのデータセットにおいて提案手法が比較手法よりも高い精度を達成しており, データセットに関わらず, 精度を維持できることがわかる. これは, 提案手法が PIA スコアにおいて prize と辺コストを正規化しているため, データセットとして用いたグラフのサイズに影響されないからだと考えられる.

5.3 実験 2

実験 2 では提案アルゴリズムの構成要素がそれぞれ, 精度にどの程度寄与するかを確認するため, 提案手法の構成要素を取り除いた場合の実験を行った. 具体的には, 初期解の選択において単純に各グループごとに prize が最大のノードを選んだもの, PIA スコアを使わず prize 順で探索候補集合 B を選んだもの, 再初期化を実行しなかったものの計 3 種類を提案手法と比較する. 実験結果を表 4 に示す. それぞれの値は 128 回の実験の結果を平均したものである.

手法	平均順位	勝率
提案手法 w/o 初期化	2.16	52.9%
提案手法 w/o PIA	1.67	69.9%
提案手法 w/o 再初期化	1.66	72.1%
提案手法	1.59	74.8%

表 4: Pubmed における ablation 実験の結果. 最も精度の高いものを太字にしている.

表 4 から提案手法の構成要素それぞれが解の精度向上に貢献していることがわかる. 特に, 初期解選択を工夫することで精度が大きく向上しており, 局所探索における初期解選択の重要性が確認できる.

5.4 実験 3

実験 3 では, ハイパーパラメータが解の傾向にどのような影響を与えるかを確認するため, スケール係数 λ , グループ数 m , 劣モジュラ関数 f それぞれを変更して実験を行った. 実験結果を表 5, 表 6, 表 7 に示す. それぞれの値は 128 回の実験の結果を平均したものである.

手法	$\lambda = 0.5$	$\lambda = 1.0$	$\lambda = 2.0$	$\lambda = 4.0$
Max-Prize	4.46	4.09	3.92	3.68
pcst_fast	4.51	5.13	5.52	5.73
exENSteiner	2.86	2.85	2.61	2.57
FastAPP	4.94	4.86	4.74	4.63
ImprovAPP	2.80	2.63	2.57	2.57
提案手法	1.43	1.46	1.65	1.83

表 5: Pubmed におけるスケール係数 λ 別の平均順位. 最も精度の高いものを太字にしている.

手法	$m = 4$	$m = 8$	$m = 12$	$m = 16$
Max-Prize	4.11	3.99	4.14	3.91
pcst_fast	5.02	5.08	5.40	5.38
exENSteiner	2.63	2.93	2.62	2.71
FastAPP	4.70	4.90	4.70	4.86
ImprovAPP	2.86	2.63	2.48	2.59
提案手法	1.68	1.48	1.67	1.55

表 6: Pubmed におけるグループ数 m 別の平均順位. 最も精度の高いものを太字にしている.

手法	log	max	$sqrt$	$topk$
Max-Prize	4.96	3.45	4.17	3.57
pcst_fast	3.43	6.00	5.45	6.00
exENSteiner	3.22	1.94	2.67	3.06
FastAPP	5.30	4.83	4.98	4.05
ImprovAPP	2.78	1.85	2.62	3.31
提案手法	1.31	2.93	1.13	1.00

表 7: Pubmed における劣モジュラ関数 f 別の平均順位. 最も精度の高いものを太字にしている.

表 5, 表 6 はそれぞれスケール係数, グループ数を変更した場合の既存手法と提案手法の平均順位を示す. それら表から, 提案手法がスケール係数, グループ数に関わらず一貫して高い精度を発揮していることがわかる. 表 7 は劣モジュラ関数を変更した場合の実験結果を示す. $topk$ 関数では $k = 3$ とした. 表 7 では f として max 関数を用いた場合に既存手法に比べて精度が低いことが確認できる. これは, max 関数が限界通減効果の極端に大きい関数であり, 一つのグループに対して複数のノードを解のサブグラフに加えることが目的関数値を悪化させるが, 提案アルゴリズムではそのような極端な限界通減効果に対応することが難しいからだと考えられる. 同じく限界通減効果が比較的強い関数である log 関数を使った場合にも, $sqrt$ や $topk$ に比べて精度が低くなっている.

6 結 論

本研究では, グループ被覆を満たしつつ, 各グループ内で高適合度のノードを優先して選択する根拠サブグラフ抽出問題を定式化, およびその問題に対する近似解法を提案し, その精度を検証した. 提案問題は, 劣モジュラ関数を導入することで, グループごとの情報の網羅度を定量的に扱い, 限界通減効果を目的関数として表現した. また, 提案アルゴリズムは, 有望な初期解の選択, PIA スコアによる探索空間の絞り込み, 再初期化によって高い精度での根拠サブグラフ探索を実現した.

今後の展望としては, 再初期化前の局所探索の簡素化や出力におけるサブグラフ構築の高度化に取り組み, より高速に高精度な根拠サブグラフ抽出ができるよう, 提案アルゴリズムの改善および拡張に取り組む予定である.

謝 辞

本研究は JSPS 科研費 JP25H01117 の助成を受けたものです.

文 献

- [1] Dennis Diefenbach, Vanessa Lopez, Kamal Singh, and Pierre Maret. Core techniques of question answering systems over knowledge bases: A survey. *Knowledge and Information Systems*, 55(3), 2018.
- [2] Arnaldo Pereira, Alina Trifan, Rui Pedro Lopes, and José Luís Oliveira. Systematic review of question answering over knowledge bases. *IET Software*, 16(1):1–13, 2022.
- [3] Ridho Reinanda, Edgar Meij, and Maarten de Rijke. Knowledge graphs: An information retrieval perspective. *Foundations and Trends® in Information Retrieval*, 14(4):289–444, 2020.
- [4] Zhenghao Liu, Chenyan Xiong, Maosong Sun, and Zhiyuan Liu. Entity-duet neural ranking: Understanding the role of knowledge graph semantics in neural information retrieval. In *ACL*, 2018.
- [5] Qijun Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. A survey on knowledge graph-based recommender systems. *TKDE*, 34(8):3549–3568, 2022.
- [6] Ying Huang et al. A knowledge graph based question answering method for medical domain. *PeerJ Computer Science*, 7:e667, 2021.
- [7] Ruby Rani, Mahender Kumar, Gregory Epiphaniou, and Carsten Maple. ICSThreatQA: A knowledge-graph enhanced question answering model for industrial control system threat intelligence. *Expert Systems with Applications*, 301:130180, 2026.
- [8] Francesco Sovrano, Monica Palmirani, and Fabio Vitali. Legal knowledge extraction for knowledge graph based question-answering. In *JURIX*, pages 143–153, 2020.
- [9] Lukas Bahr, Christoph Wehner, Judith Wewerka, José Bittencourt, Ute Schmid, and Rüdiger Daub. Knowledge graph enhanced retrieval-augmented generation for failure mode and effects analysis. *Journal of Industrial Information Integration*, 45:100807, 2025.
- [10] Fan Yang et al. Keyword search on large graphs: A survey. *Data Science and Engineering*, 2021.
- [11] Xiaoxin He, Yijun Tian, Yifei Sun, Nitesh V. Chawla, Thomas Laurent, Yann LeCun, Xavier Bresson, and Bryan Hooi. G-Retriever: Retrieval-augmented generation for textual graph understanding and question answering. In *NeurIPS*, 2024.
- [12] Yunshi Lan, Jing Jiang, Xin Jiang, Wayne Xin Zhao Wang, and Ji-Rong Wen. A survey on complex knowledge base question answering: Methods, challenges and solutions. In *IJCAI*, pages 4483–4491, 2021.
- [13] Xiaolu Lu, Soumajit Pramanik, Rishiraj Saha Roy, Abdalghani Abujabal, Yafang Wang, and Gerhard Weikum. Answering complex questions by joining multi-document evidence with quasi knowledge graphs. In *SIGIR*, pages 105–114, 2019.
- [14] Wenjie Li, Yanan Qin, Jeffrey Xu Yu, and Rong Mao. Efficient and progressive group steiner tree search. In *SIGMOD*, 2016.
- [15] Aaron Archer, MohammadHossein Bateni, Mohammad-Taghi Hajiaghayi, and Howard Karloff. Improved approximation algorithms for prize-collecting Steiner tree and TSP. *SIAM Journal on Computing*, 40(2):309–332, 2011.
- [16] Naveen Garg, Goran Konjevod, and R. Ravi. A polylogarithmic approximation algorithm for the group steiner tree

- problem. *Journal of Algorithms*, 37(1):66–84, 2000.
- [17] Michel X. Goemans and David P. Williamson. A general approximation technique for constrained forest problems. *SIAM Journal on Computing*, 24(2):296–317, 1995.
- [18] David S. Johnson, Maria Minkoff, and Steven Phillips. The prize-collecting steiner tree problem: Theory and practice. In *SODA*, pages 760–769, 2000.
- [19] Yahui Sun, Xiaokui Xiao, Bin Cui, Saman K. Halgamuge, Theodoros Lappas, and Jun Luo. Finding group steiner trees in graphs with both vertex and edge weights. *Proc. VLDB Endow.*, 14(7):1137–1149, 2021.
- [20] Jaime Carbonell and Jade Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *SIGIR*, pages 335–336, 1998.
- [21] Bolin Ding, Jeffrey Xu Yu, Shan Wang, Lu Qin, Xiao Zhang, and Xuemin Lin. Finding top-k min-cost connected trees in databases. In *ICDE*, pages 836–845, 2007.
- [22] Yunliang Shi, Jinhua Zhuang, Yu Zhang, Qifan Wang, Changyou Chen, Yaliang Li, Jun Yuan, and Jiawei Han. Keyword-based knowledge graph exploration based on quadratic group steiner trees. In *IJCAI*, 2021.
- [23] Soumajit Pramanik, Jesujoba Alabi, Rishiraj Saha Roy, and Gerhard Weikum. Uniqorn: Unified question answering over RDF knowledge graphs and natural language text. *Journal of Web Semantics*, 83:100833, 2024.
- [24] Nicholas Thomas Walker, Pierre Lison, Laetitia Hilgendorf, Nicolas Wagner, and Stefan Ultes. Retrieving relevant knowledge subgraphs for task-oriented dialogue. In *SIGDIAL*, pages 513–526, 2025.
- [25] Z. Ali, A. Haldar, K. Korhonen, and J. Kontio. Pythia-rag: Retrieval-augmented generation with unified multimodal knowledge graph. *Knowledge-Based Systems*, 335:115200, 2026.
- [26] Mehdi Kargar, Aijun An, and Morteza Zihayat. Efficient bi-objective team formation in social networks. In *Machine Learning and Knowledge Discovery in Databases*, pages 483–498, 2012.
- [27] Xiang Wei, Wei Lu, and Weiwei Xing. A rapid multi-source shortest path algorithm for interactive image segmentation. *Multimedia Tools and Applications*, pages 21547–21563, 2017.
- [28] Kurt Mehlhorn. A faster approximation algorithm for the Steiner problem in graphs. *Information Processing Letters*, 27(3):125–128, 1988.
- [29] Zhilin Yang, William W. Cohen, and Ruslan Salakhutdinov. Revisiting semi-supervised learning with graph embeddings. In *ICML*, pages 40–48, 2016.
- [30] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.
- [31] Chinmay Hegde, Piotr Indyk, and Ludwig Schmidt. A nearly-linear time framework for graph-structured sparsity. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, volume 37, pages 928–937, 2015.
- [32] Aida Rahmattalabi, Shahin Jabbari, Himabindu Lakkaraju, Phebe Vayanos, Max Izenberg, Ryan Brown, Eric Rice, and Milind Tambe. Fair influence maximization: A welfare optimization approach. In *AAAI*, pages 11630–11638, 2021.
- [33] Hui Lin and Jeff Bilmes. A class of submodular functions for document summarization. In *ACL*, pages 510–520, 2011.

GLCSRec: Integrating Graph Collaborative Signals with Large Language Models for Item Cold-Start Recommendation

Ying LIN[†], Chongxian CHEN[†], Xin FAN[†], and Hayato YAMANA^{††}

[†] Graduate School of Fundamental Science and Engineering,
3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan

^{††} Faculty of Science and Engineering, Waseda University,
3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan

E-mail: [†]alisa@ruri.waseda.jp, ^{††}chenc@toki.waseda.jp, ^{†††}fan_xin@fuji.waseda.jp,
^{††††}yamana@yama.info.waseda.ac.jp

Abstract Recommendation systems aim to precisely match users with items. However, newly introduced items often encounter the item cold-start problem due to insufficient historical interaction data. This data sparsity undermines the performance of collaborative filtering methods. Although large language models (LLMs) offer potential for processing new item content through their semantic reasoning capabilities, they inherently lack integration of collaborative signals, which constitutes a critical structural limitation. To address this, we propose GLCSRec (Graph-LLM Cold-Start Recommender). Unlike previous approaches that directly perform vector space alignment, this model utilizes LightGCN to extract user features from their historical interactions and projects them into soft prompts through a multilayer perceptron (MLP). The soft prompts consist of continuous and trainable vectors and are embedded in the input layer of the LLM, thereby guiding the causal attention mechanism of the LLM to conduct personalized semantic reasoning and preference prediction for the metadata of cold-start items. The model is trained with the Bayesian personalized ranking (BPR) loss, with optimization applied exclusively to the user embeddings and the projector parameters. A comprehensive experimental evaluation demonstrated the effectiveness of GLCSRec: (1) Compared to the representative baseline models, the Recall@10 increased from 0.0534 to 0.0765. (2) The ablation experiments confirmed the necessity of the collaborative signal, as replacing the GNN embeddings with semantic features led to a decrease in Recall@10 from 0.0805 to 0.0545. (3) The sensitivity and robustness analysis verified the model’s stability, maintaining MRR@10 stable at 0.3484 even as data sparsity increased.

Key words Recommender Systems, Item Cold-Start, Graph Neural Networks (GNN), Large Language Models (LLM), Soft Prompt

1 Introduction

Recommender systems play a vital role in modern web services, acting as the primary engine for information filtering [1]. Collaborative filtering models [2], such as LightGCN [3], are academic benchmark models due to their exceptional ability to capture high-order user-item interaction structures. However, these models fundamentally rely on the density of the interaction graph, leading to the item cold-start problem [4, 5]. Specifically, in the item cold-start scenario where new items in the test set have never appeared in the training set, traditional models fail to recommend due to they lack historical data. Our experiments show that in this scenario, LightGCN’s Recall@10 drops to 0.0, highlighting the need to develop models that understand the content of cold-start items to recommend.

Large language models (LLMs) [6] possess semantic reasoning capabilities and have the potential to provide semantic reasoning for cold-start items. However, our experiments reveal that using LLMs alone results in poor performance (Recall@10 0.0067) due to there is a lack of collaborative signals from the user history; although LLMs can understand the cold-start items, they are unable to identify which users will interact with them [6].

How to effectively integrate the highly structured user-item relationships captured by graph neural networks (GNNs) [7] into the semantic reasoning capabilities of LLMs is a challenge. To bridge this gap, we propose the GLCSRec framework, which integrates the collaborative signals of GNNs and the semantic reasoning capabilities of LLMs. Instead of using a direct alignment method, we project the user fea-

tures extracted by LightGCN from the user interaction history through a multilayer perceptron (MLP) into the soft prompts [8]. The resulting collaborative signals are transformed into continuous vectors, used as soft prompts [9]. These prompts can effectively guide the LLM to enable the causal attention mechanism to perform personalized semantic reasoning and preference prediction for the metadata of cold-start item, thereby compensating for the lack of interaction data.

The contributions of this thesis are as follows:

1. Our proposed method effectively integrates the structured signals of GNNs into the semantic reasoning capabilities of LLMs, rather than only relying on the content’s metadata. Compared with the representative baselines, it has increased the Recall@10 from 0.0534 to 0.0765.
2. The ablation experiment results show that after removing the GNN signal and only relying on the textual description, the Recall@10 drops by 32.3%. Therefore, we have confirmed that GNN can effectively capture the collaborative signals that LLM cannot derive, proving the irreplaceability of structural signals.
3. We verified the GLCSRec that even when the training interaction data is reduced by 46.7% (the cold-start ratio increases from 0.1 to 0.5), the MRR@10 still remains stable at 0.3484. This indicates that our architecture effectively integrates the GNN signal through the soft prompt to the LLM, and can maintain precise semantic reasoning and ranking even in sparse scenarios.

2 Related Work

This chapter reviews related work, addressing the cold-start problem, by classifying them into 1) content-based and hybrid models, 2) graph neural networks-based, and 3) large language models-based. By analyzing the limitations of related work, we identify the research gaps that motivate our proposed framework.

2.1 Content-based and Hybrid Models

Content-based filtering [10] addresses item cold-start by substituting interaction data with metadata. SBERT [11] encodes textual descriptions to calculate similarity with user preference [12]. While Ding et al. [13] utilized semantic retrieval, they focus on explicit relevance, overlooking high-order collaborative signals.

Hybrid models like DeepFM [14] rely on trained ID embeddings, failing for cold-start items lacking vectors. Other models [1,15] struggle to balance metadata and collaborative signals [5], potentially separating reasoning from ranking [6].

Alignment approaches learn from ID-based space to semantics [15]: MeLU [16] uses few-shot learning but requires

initial signals; CLCRec [17] utilizes contrastive objectives but may lose information during vector mapping.

2.2 Graph Reconstruction for Cold-Start Recommendation

Graph neural networks (GNNs) model user-item interactions as bipartite graphs [7, 18], utilizing topological structure to capture high-order connectivity. Representative models like NGCF [18] and LightGCN [3] focus on collaborative smoothing via neighborhood aggregation; however, standard GNNs remain inherently transductive [19, 20], relying on edges to propagate signals and failing to generate representations for isolated cold-start nodes. To overcome this, CGRC [21] explores graph reconstruction by constructing edges from content features to alleviate isolation. This approach remains limited as surface-level similarity introduces structural noise [22], weakening user representation robustness. Furthermore, treating text as construction features rather than semantic guidance [21] lacks deep reasoning, failing to understand subtle preferences and text descriptions.

2.3 LLMs for Cold-start Recommendation

LLMs enable generative semantic reasoning [6, 23], interpreting complex intentions and contexts where traditional models struggle. In Zero-Shot settings, LLMs (e.g., TinyLlama [24]) rank preferences but encounter collaborative gaps [6, 25] from unrecognized historical behavior and popularity bias, failing to capture personalized requirements [25]. TALLRec [26] and RecLM [27] utilize lightweight fine-tuning or instruction tuning to integrate behavioral representations, yet lengthy natural language prompts increase computational demands and linear text limits capturing high-order collaborative relationships. Soft prompt methods [9, 28] address costs by freezing backbones and optimizing continuous virtual tokens. While GraphPrompter [8] encoded structural information for general graph learning, it remains unadapted to integrate collaborative signals into cold-start item inference lacking interaction history.

Table 1: Summary of representative related work

Category	Models	Overview	Research Gaps
Content & Hybrid	DeepFM [14], CLCRec [17]	Metadata encoding; contrastive alignment.	Information loss; difficulty in balancing signals.
GNN-based	LightGCN [3], CGRC [21]	Capture high-order connectivity; reconstructing graph.	Isolation of cold-start nodes; structural noise.
LLM-based	TALLRec [26], RecLM [27]	Generative reasoning; instruction tuning.	Lack of collaborative signals; high costs.

3 Preliminaries

This chapter explains the theoretical foundations and

mathematical notations required for our proposed GLCSRec framework.

3.1 Bayesian Personalized Ranking (BPR)

We adopt a pairwise optimization strategy, i.e., Bayesian personalized ranking (BPR) [29]. This method learns personalized rankings based on pairs of items, which assumes that the user’s preference for observed items is higher than unobserved items. This optimization is robust for ranking tasks and helps maintain recommendation stability even in data-sparse scenarios. The BPR loss (L_{BPR}) is defined:

$$L_{BPR} = - \sum_{(u, i^+, i^-) \in D} \ln \sigma(\hat{y}_{ui^+} - \hat{y}_{ui^-}) + \lambda_{\theta} \|\Theta\|_2^2 \quad (1)$$

3.2 Graph Neural Networks (GNNs)

LightGCN simplifies message passing using normalized adjacency matrix $\tilde{\mathbf{A}} = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$, updating embeddings: $\mathbf{E}^{(k+1)} = \tilde{\mathbf{A}} \mathbf{E}^{(k)}$. However, standard GNNs are limited by transductive nature [20]; for isolated cold-start nodes $i \in I_{cold}$ with no interaction history, the propagation rule becomes ineffective:

$$\mathbf{e}_i^{(k+1)} = \tilde{\mathbf{A}}_{i,:} \mathbf{E}^{(k)} = \mathbf{0} \quad (2)$$

3.3 Parameter-efficient Fine-tuning (PEFT)

To address computational costs, soft prompt methods [9] freeze backbone architectures and optimize continuous virtual tokens $\mathbf{P} = \{p_1, \dots, p_l\}$. Following GraphPrompter [8], GNNs encode local structures into vectors projected via MLP:

$$\tilde{\mathbf{X}}_i = \text{MLP}(\text{GNN}(\mathbf{G}_{s_i})) \in \mathbb{R}^{d_l} \quad (3)$$

The final LLM input concatenates structural soft prompts and text embeddings \mathbf{T}_{emb} :

$$\mathbf{H}_{input} = [\tilde{\mathbf{X}}_i; \mathbf{T}_{emb}] \quad (4)$$

This allows utilizing structural signals and semantic knowledge without complete parameter fine-tuning.

4 The Proposed Method: GLCSRec

The design of GLCSRec (Graph-LLM Cold-Start Recommender) is motivated by the need to bridge the information gap between structured collaborative signals and semantic text knowledge in the item cold-start recommendation task. We propose a collaborative prompt mechanism and use a parameter-efficient strategy to fine-tune the complete model. The overall architecture is shown in Figure 1.

4.1 Preparation of Data

Cold-Start Item Partitioning. Let $U = \{u_1, \dots, u_m\}$ and $I = \{i_1, \dots, i_n\}$ denote users and items, where interactions are represented by binary matrix $\mathbf{Y} \in \{0, 1\}^{M \times N}$. The item set is partitioned into disjoint subsets: $I = I_{train} \cup I_{cold}$, where $I_{train} \cap I_{cold} = \emptyset$. I_{train} contains history for structural learning, while I_{cold} denotes new items lacking history. The objective is to learn mapping function $f : (U, I_{cold}) \rightarrow \mathbb{R}$ predicting preference score \hat{y}_{uj} for cold-start item $j \in I_{cold}$, relying on metadata \mathbf{T}_j and user collaborative context.

Preparation of GNN. We implement a masking protocol, removing all interactions for $j \in I_{cold}$ from the bipartite graph to eliminate transductive leakage. Since item j has no neighboring nodes, the messaging mechanism fails to propagate cooperative signals, resulting in decoupled representation:

$$\mathbf{e}_j = \alpha_0 \mathbf{e}_j^{(0)} + \sum_{k=1}^K \alpha_k \mathbf{0} = \alpha_0 \mathbf{e}_j^{(0)} \quad (5)$$

Item ID embedding \mathbf{e}_j remains as unoptimized random noise from random initialization. Lacking interaction data, these embeddings cannot be updated during training, lead-



Figure 1: The overall architecture of GLCSRec. The process begins with a pre-trained LightGCN (Left) for extracting high-order user embeddings, followed by a multi-layer projector (Middle) that aligns these signals with the LLM input space. Finally, the Soft Prompts (Right) act as personalized instructions to guide the frozen LLM for item cold-start recommendation.

ing to failure in capturing user preferences. This forces the model to turn to semantic space, relying on cold-start item’s metadata \mathbf{T}_j (e.g., title and genres) to perform inductive reasoning through the alignment interface of the frozen LLM.

4.2 Collaborative Signal Encoder

We adopt LightGCN [3] to extract high-order user preference patterns from the historical interaction graph [7], acting as a low-pass filter to improve embedding quality in sparse graphs. Following standard settings [3], we set layer depth $K = 3$ to capture collaborative filtering while avoiding over-smoothing, adopting a layer fusion strategy to aggregate information. The final collaborative user representation \mathbf{e}_u is obtained by $\mathbf{e}_u = \sum_{k=0}^K \alpha_k \mathbf{e}_u^{(k)}$, where $\mathbf{e}_u^{(k)}$ denotes the embedding vector at layer k and $\alpha_k = \frac{1}{K+1}$ denotes the mean pooling fusion weight. This uniform weighting strategy prevents over-smoothing while capturing high-order connectivity, allowing the final representation to capture local and global signals.

Two-Stage Training Strategy. To optimize efficiency and collaborative signal stability, we propose a decoupled two-stage training strategy. In the first stage, the GNN encoder is pre-trained independently on the observed user-item interaction graph to ensure initial user embeddings \mathbf{e}_u are embedded with collaborative signals. We optimize GNN parameters using Bayesian personalized ranking (BPR) [29] to optimize pairwise ranking, making the structural foundation reliable. In the second stage, we import pre-trained user embedding vectors into the semantic alignment interface. To ensure high-order collaborative signals are not damaged, we froze the GNN propagation structure and LLM backbone parameters, optimizing only for the user embedding vector and alignment parameters. Through the soft prompt mechanism, gradients are transmitted while maintaining the LLM backbone frozen, bridging the modality gap and enabling collaborative signals to precisely map to the semantic space, complying with parameter efficiency fine-tuning (PEFT) [6, 9].

4.3 Semantic Alignment Interface

We use a multi-layer projection function $\phi(\cdot)$ composed of MLP to perform cross-modal transformation, bridging the modality gap between d -dimensional collaborative latent space and the frozen LLM high-dimensional embedding space. This multi-layer structure allows for non-linear alignment, mapping user-collaborative embedding vector $\mathbf{e}_u \in \mathbb{R}^d$ to LLM token space $\mathbb{R}^{L \times D_{LLM}}$ as defined: $\phi(\mathbf{e}_u) = \mathbf{W}_2 \cdot \sigma(\mathbf{W}_1 \cdot \mathbf{e}_u + \mathbf{b}_1) + \mathbf{b}_2$.

The transformation employs weight matrices $\mathbf{W}_1 \in \mathbb{R}^{d_h \times d}$, $\mathbf{W}_2 \in \mathbb{R}^{(L \cdot D_{LLM}) \times d_h}$, bias vectors $\mathbf{b}_1, \mathbf{b}_2$, and non-linear activation function $\sigma(\cdot)$ (ReLU) to map input \mathbf{e}_u ($d = 64$) through latent space ($d_h = 256$) to intermediate features; these are reshaped into $L = 32$ continuous virtual tokens to

align with LLM embedding dimension D_{LLM} . Collaborative signals are refined into soft prompts $\mathbf{P}_u \in \mathbb{R}^{L \times D_{LLM}}$, acting as personalized instructions that encode interaction history in semantic latent space [9, 30].

4.4 LLM-Integrated Recommendation

GLCSRec utilizes extensive world knowledge in pre-trained LLMs to conduct inductive inferences on cold-start items, bridging the collaborative gap. For cold-start item $j \in I_{cold}$, we retrieve textual metadata and convert it into token embeddings $\mathbf{E}_{item} \in \mathbb{R}^{T \times D_{LLM}}$. We adopt a prefix injection strategy, attaching soft prompt \mathbf{P}_u before the semantic embedding sequence \mathbf{E}_{item} to construct the unified input matrix $\mathbf{X}_{in} = [\mathbf{P}_u; \mathbf{E}_{item}] \in \mathbb{R}^{(L+T) \times D_{LLM}}$, where $[\cdot; \cdot]$ denotes concatenation along the sequence length dimension.

\mathbf{X}_{in} conditions the causal attention mechanism, allowing the transformer backbone to attend back to virtual collaborative tokens while encoding item metadata, ensuring semantic reasoning is grounded in interaction history. Utilizing TinyLlama-1.1B [24] as the frozen backbone, we observe the hidden state of the final token $\mathbf{h}_{last} \in \mathbb{R}^{D_{LLM}}$ to extract relevance score $\hat{y}_{ui} = \mathbf{w}_{score}^\top \mathbf{h}_{last} + b_{score}$ via a linear score header. Trainable parameters $\mathbf{w}_{score} \in \mathbb{R}^{D_{LLM}}$ and $b_{score} \in \mathbb{R}$ are optimized using BPR loss [29] to maximize the margin between positive items and negative samples, mitigating computational costs by restricting updates to projectors and scoring heads.

4.5 Optimization Strategy

To reduce computational costs, we adopt parametric efficient fine-tuning (PEFT), freezing transformer backbone Θ_{LLM} and GNN propagation weights while limiting the trainable parameter set to $\Theta_{train} = \{\mathbf{E}_u, \Theta_\phi, \mathbf{w}_{score}, b_{score}\}$, where \mathbf{E}_u denotes learnable user embedding vectors, Θ_ϕ denotes soft prompt projector parameters, and $\mathbf{w}_{score}, b_{score}$ denote scoring head parameters. We employ BPR loss [29] augmented with a regularization term: $L = L_{BPR}(\Theta_{train}) + \lambda \|\Theta_{train}\|^2$.

When Θ_{LLM} is frozen, gradients backpropagate through transformer layers to update the projector and user embeddings, ensuring end-to-end alignment between collaborative signals and semantic reasoning space [9]. PEFT enables memory efficiency by eliminating the necessity of storing gradients for the massive backbone, reducing memory cost from $O(\Theta_{LLM})$ to $O(\Theta_{train})$, where Θ_{train} constitutes $\approx 1.61\%$ of total parameters. Freezing Θ_{LLM} enhances training efficiency by eliminating computational overhead for calculating backbone weight gradients, limiting optimization complexity to $O(\Theta_{train})$, as detailed in Table 2.

5 Experiments

This chapter explains the experimental evaluation frame-

Table 2: Parameter Configuration and Complexity Analysis (M denotes Million).

Component	Notation	Params (M)	Status
LLM Backbone	Θ_{LLM}	1,100	Frozen
User Embeddings	\mathbf{E}_u	0.39	Trainable
Projector	Θ_ϕ	17.36	Trainable
Scoring Head	$\mathbf{w}_{score}, b_{score}$	0.002	Trainable
Metric	Full Fine-Tuning	GLCSRec (Ours)	
Trainable Params	$\approx 1,118$ M	≈ 17.75 M (1.61%)	
Space Complexity	$O(\Theta_{LLM} + \Theta_{train})$	$O(\Theta_{train})$	
Time Complexity	$O(\Theta_{LLM} + \Theta_{train})$	$O(\Theta_{train})$	

work to confirm the effectiveness of the GLCSRec model. We will evaluate the performance of this framework in cold-start scenarios, where traditional collaborative filtering methods fail due to the lack of historical interaction data for new items. We recall the following RQs:

- **RQ1 (Overall Performance):** Does GLCSRec perform better than the baseline models in recommendation accuracy for cold-start items?
- **RQ2 (Ablation Study):** How much does the GNN-based collaborative signal contribute to the performance?
- **RQ3 (Robustness & Sensitivity):** How robust is the model to data sparsity, and how do hyperparameters like soft prompt length affect performance?

5.1 Experimental Settings

We utilize the widely used MovieLens-1M dataset. To ensure the quality of the graph structure and reduce interference from inactive nodes, we apply a 5-core filtering strategy, which removes users and items with fewer than 5 interactions. After preprocessing, the final dataset consists of 6,040 users and 3,706 active items, with a total of 1,000,209 interactions, yielding an interaction density of 4.47%.

Cold-Start Protocol. To simulate a realistic cold-start scenario, we adopt a item-side evaluation protocol:

1. Item split: Following the strategy in CGRC [21], we randomly select 70% of the items (2,593 items) to form the training set (warm items). The remaining 30% (1,113 items) are cold items. We split the interactions of these cold items into two parts: 15% for validation and 15% for testing. Because some items have very few interactions, they might not show up in one of the sets. That is why the actual item counts are 1,080 for validation and 1,087 for testing. The training set has 683,826 interactions, with a density of 4.37%.
2. Interaction masking: We masked interactions for all cold items in the validation and test sets ($y_{ui} = 0$) during

training. This ensures that the GNN module cannot learn any structural ID embeddings for these items, forcing the model to rely on inductive reasoning via the LLM backbone.

3. Feature extraction: We employed a pre-trained SBERT model [11] to encode item metadata (i.e., titles and genres) into 384-dimensional vectors, serving the semantic context for the LLM.

Baselines. To evaluate the effectiveness of GLCSRec, we prepare reference models and cold-start baseline models, as described below:

1. Reference Models. We include two reference models to quantify the structural limitations of GNN and LLM in cold-start settings, the following models are not specifically designed to address the cold-start problem:

- **LightGCN [3]:** It is a representative collaborative filtering (CF) model. We set the embedding size to 64, learning rate to 1×10^{-3} , and batch size to 4096. Following the standard configuration in the RecBole framework, the model is trained for 30 epochs. Since LightGCN relies on learned ID embeddings, it is unable to generalize to unseen items, validating the failure of traditional CF under cold-start settings.
- **Zero-Shot LLM:** It is a content-only baseline utilizing the frozen TinyLlama-1.1B-Chat model [24]. The model is evaluated in half-precision (float16) without parameter updates. User preference is predicted via prompt-based inference (e.g., answer yes or no), where the probability of the affirmative response serves as the ranking score. This baseline represents the capability of pure LLM reasoning; however, it lacks collaborative signals from the user-item interaction graph, resulting in a "collaborative gap" that limits its personalization.

2. Cold-Start Baselines. These models are specifically designed to address the item cold-start problem:

- **Cosine Similarity:** It is a classic non-parametric content-based baseline. We employ the SBERT model [11] to encode item textual descriptions into 384-dimensional semantic vectors. User profiles are constructed by averaging the embeddings of their historical interactions. Recommendation scores are computed using cosine similarity between user and item representations.
- **Hybrid GNN:** It is a representative hybrid alignment baseline, inspired by hybrid recommendation models that combine collaborative and content signals [15, 31]. Specifically, user embeddings from GNN encoder are projected into the same semantic space as item text embeddings via a two-layer MLP ($64 \rightarrow 128 \rightarrow 384$) with ReLU activation function. The model is trained for 20

epochs using a learning rate of 1×10^{-3} and a cosine embedding loss. This baseline reflects a commonly adopted alignment paradigm that combines collaborative signals with content semantics.

- CLCRec [17]: It is a competitive cold-start recommendation method based on contrastive learning. We adopt the hyperparameters following the configurations suggested by the original authors, setting the learning rate to 0.001, regularization weight to 0.1, temperature $\tau = 2.0$, and contrastive loss weight $\lambda = 0.5$. An early stopping strategy based on validation recall is followed to select the best-performing checkpoint.

Implementation Details. We implemented GLCSRec using PyTorch 2.4.1 with Python 3.10. All model were trained on a single NVIDIA Tesla V100 GPU environment supporting CUDA 12.1. We adopt TinyLlama-1.1B as the frozen LLM backbone. For the collaborative encoder, the LightGCN embedding dimension is set to 64. The projector is a two-layer MLP with a hidden dimension of 256. As for the input settings, we set the soft prompt length to $L = 32$ and the maximum text sequence length to 128 tokens. The model is trained for 5 epochs using the Adam optimizer [32] with a learning rate of $1e - 4$ and a batch size of 32.

5.2 Overall Performance (RQ1)

Table 3 summarizes the ranking performance comparison between GLCSRec and the baselines.

Table 3: Overall ranking performance comparison for cold-start items only. All models are evaluated using a fixed random seed (seed=42) to ensure identical data splitting.

Model	Recall@10	NDCG@10	Hit@10	MRR@10
LightGCN	0.0000	0.0000	0.0000	0.0000
Zero-Shot LLM	0.0067	0.0192	0.1675	0.0433
Cosine Similarity	0.0534	0.1093	0.5445	0.2449
Hybrid GNN	0.0463	0.1035	0.5156	0.2377
CLCRec	0.0200	0.0522	0.3156	0.1309
GLCSRec (proposed)	0.0805	0.1724	0.7058	0.3484

Performance Analysis. The results reveal several critical insights: The results reveal the following insights:

- LightGCN fails completely (metrics ≈ 0), confirming that pure GNNs cannot handle unseen nodes without learned ID embeddings, making it incapable of recommending cold-start items.
- Zero-shot LLM performs poorly (Recall 0.0067), worse than content-based methods. This indicates that the world knowledge of LLMs is insufficient to capture personalized cold-start recommendations.
- Cosine similarity is a content-based model, outperforms other models (Recall ≈ 0.0534), its direct semantic comparison strategy can reduce the risk of underfitting or

overfitting in the cold-start setting.

- Hybrid GNN and CLCRec rely on interaction structures to align their latent spaces. In the cold-start setting, the mapping functions learned by these models on popular items cannot be effectively generalized to isolated nodes (cold-start item). This result shows the difficulty of mapping the semantic space to the collaborative space.
- GLCSRec achieves the best performance across all metrics. By integrating GNN-aggregated collaborative signals as soft prompts, our model successfully bridges the gap between collaborative signals and semantic reasoning for cold-start items.

5.3 Ablation Study (RQ2)

To validate the effectiveness of our core architectural choice, we analyze the contribution of the GNN-based collaborative signal.

Table 4: Ablation study on the collaborative signal. The variant "collaborative vs. semantic" refers to replacing structural GNN embeddings with semantic user profiles aggregated from item history.

Variant	Recall@10	NDCG@10	Hit@10	MRR@10
GLCSRec: Full	0.0805	0.1724	0.7058	0.3484
GLCSRec: Semantic-only	0.0545	0.1233	0.5798	0.2695

Impact of Collaborative Signal (collaborative vs. semantic). In our framework, the projector maps high-order user embeddings from LightGCN into the LLM's prompt space. To verify the necessity of these structural signals, we evaluated a variant, collaborative vs. semantic, where the GNN-based embeddings were replaced with semantic user profiles. These profiles were generated by averaging the embedding vectors of users' historical items encoded by LLM, representing a semantic of user preference without collaborative signal.

As shown in Table 4, removing the graph collaborative signal leads to a performance drop. This confirms that the LLM cannot effectively recommend items relying on historical semantic context. It requires collaborative signals from the GNN to capture interaction history that are not present in item descriptions, achieving more accurate cold-start recommendations.

5.4 Sensitivity and Robustness Analysis (RQ3)

5.4.1 Hyperparameter Sensitivity: Prompt Length

We explore the impact of the soft prompt length $L \in \{16, 32, 64\}$ on the recommendation performance of GLCSRec. The results are shown in Table 5, showing how different lengths of soft prompts affect the evaluation metrics.

Table 5: Performance sensitivity to Soft Prompt Length L (Seed 42).

Length (L)	Recall@10	NDCG@10	Hit@10	MRR@10
16	0.0808	0.1674	0.6969	0.3331
32	0.0805	0.1724	0.7058	0.3484
64	0.0806	0.1760	0.6925	0.3635

Increasing the prompt length L brings a slight improvement to the ranking precision metrics (NDCG@10, MRR@10). For instance, NDCG@10 rises from 0.1674 to 0.1760, indicating that longer virtual prompt sequences can provide LLMs with richer context for detailed item ordering. Other evaluation metrics (Recall@10 and Hit@10) perform consistently under different L settings, with a limited variation in overall performance. This indicates that the model has stability of prompt length, also shows that recommendation performance is mainly determined by the GNN structure rather than prompt length.

5.4.2 Training Stability and Convergence

Table 6 and Figure 2 summarize the computational cost and convergence of GLCSRec under different prompt lengths $L \in \{16, 32, 64\}$.

Table 6: Training convergence and time cost across different sequence lengths (L).

Config	Final Loss	Total Time	Speed (it/s)	Epoch
$L = 16$	0.0021	≈ 8.8 Hours	3.37	5
$L = 32$	0.0360	≈ 11.1 Hours	2.68	4
$L = 64$	0.3156	≈ 16.1 Hours	1.84	3* (fluctuated)

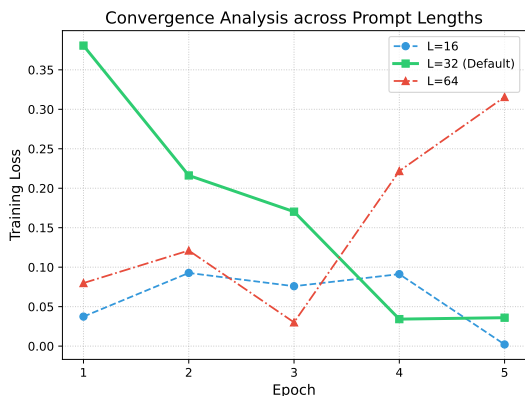


Figure 2: Training loss convergence and stability analysis over 5 epochs across different soft prompt Lengths L .

Training results indicate effective convergence; all configurations reach stable or minimal loss within 5 epochs. The default setting ($L = 32$) training loss decreases from 0.3807 to 0.0342 by the 4th epoch; negligible difference between the

4th (0.0342) and 5th epoch (0.0360) suggests stable convergence, indicating 5 epochs are sufficient. While $L = 16$ offers the highest training speed (3.37 it/s), its capacity to capture semantic collaborative signals is limited. Although the computational load of $L = 32$ increased approximately 2.3 hours more than $L = 16$, it provides stable optimization compared to $L = 64$.

The $L = 64$ shows unstable fluctuations; loss drops to 0.0301 in the 3rd epoch but rose to 0.3156 by the 5th epoch. These fluctuations, coupled with lower training speed (1.84 it/s), indicate excessively long soft prompts cause gradient noise and complicate alignment between GNN and LLM latent spaces.

5.4.3 Robustness to Random Sampling

To verify that the excellent performance of GLCSRec is not an accident of random sampling, we conducted experiments using 5 different random seeds and carried out statistical reliability tests, as shown in Table 7.

Table 7: Statistical Reliability of GLCSRec (Over 5 Random Seeds).

Metric	Recall@10	NDCG@10	Hit@10	MRR@10
Mean	0.0765	0.1666	0.6888	0.3497
Std Dev	± 0.0052	± 0.0108	± 0.0224	± 0.0241

The results show that all the evaluation metrics exhibit a high consistency. The low standard deviation (Recall@10 $\sigma = 0.0052$) indicates that the alignment of GNN collaborative signals with the LLM semantic space is robust and reproducible. These results indicate that GLCSRec can effectively capture user-item preferences and is not affected by random variations during the training process.

5.4.4 Robustness to Data Sparsity

We evaluated the robustness of GLCSRec by varying the cold-start ratio from 0.1 (Easy) to 0.5 (Hard). Table 8 summarizes the performance alongside the training data scale to provide a clear view of data scarcity.

Table 8: Robustness analysis under varying cold-start ratios. The term *train inter.* represents the total observed edges, while *density* reflects the sparsity of the training graph.

Ratio	Train Inter.	Density	Recall@10	NDCG@10	Hit@10	MRR@10
0.1 (Easy)	910,718	4.62%	0.1569	0.1521	0.5483	0.2611
0.3 (Default)	683,826	4.37%	0.0805	0.1724	0.7058	0.3484
0.5 (Hard)	485,541	4.33%	0.0471	0.1555	0.6783	0.3484

In contrast, at higher ratios (0.3 and 0.5), the increased data sparsity forces the projector to learn more robust translation rules from structural signals to semantic prompts. The MRR@10 remains consistently stable at 0.3484, indicating

that although the weakening GNN signal reduces the retrieval diversity, the stable reasoning ability of the LLM ensures that highly reliable semantic matching can rank accurately. In particular deployment environments, the cold-start ratio of items is often affected by external environmental factors, indicating the importance of this robustness.

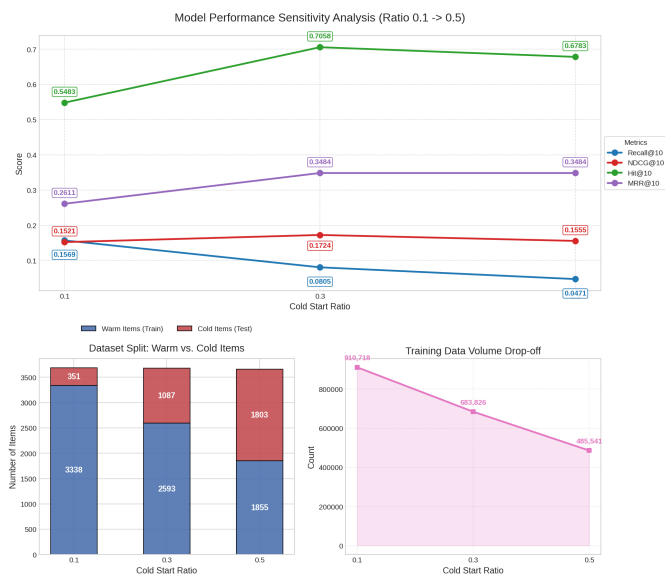


Figure 3: Model performance sensitivity analysis and data distribution between different cold-start ratios (0.1to 0.5). The above chart shows the results among various evaluation metrics as the cold-start ratio increases. The bottom-left chart shows the ratio of the warm item (training set) to the cold item (test set), and the bottom-right chart shows the decreasing trend of the total training interaction volume as the cold-start ratio changes.

6 Conclusion

In this paper, we introduced GLCSRec to address the item cold-start recommendation. This framework bridges the gap between collaborative filtering and semantic reasoning, converting the high-order interaction history captured by GNN into soft prompts. This enables the frozen LLM to inherit the knowledge of the language model while also basing the reasoning on the collaborative signals. We conduct an extensive experimental evaluation on the MovieLens-1M dataset. We demonstrated the effectiveness and robustness of the integration of GNN and LLM through soft prompt. Even when the training data was reduced by nearly 50%, the model still successfully maintain high ranking accuracy by leveraging the semantics of the backbone of LLM. This indicates that GLCSRec has achieved effective cross-modal alignment between graph structures and natural language.

Future Work. We plan to investigate knowledge distillation technology to further reduce inference delay. Addi-

tionally, integrating multi modal information (such as item images and video descriptions) can provide a more comprehensive representation for cold-start items. Finally, we aim to extend the current framework to cross-domain scenarios to assist in item recommendations in domains with extremely sparse interaction history.

References

- [1] Francesco Ricci, Lior Rokach, and Bracha Shapira. Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer, 2010.
- [2] J. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. Collaborative filtering recommender systems. In *The adaptive web: methods and strategies of web personalization*, pages 291–324. Springer, 2007.
- [3] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. LightGCN: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pages 639–648, 2020.
- [4] Jyotirmoy Gope and Sanjay Kumar Jain. A survey on solving cold start problem in recommender systems. In *2017 International Conference on Computing, Communication and Automation (ICCCA)*, pages 133–138. IEEE, 2017.
- [5] Natalija Glisovic, Danica Kragic, and Martin Tegner. Item cold start in e-commerce recommender systems: A survey. volume 13, pages 164702–164722. IEEE, 2025.
- [6] Likang Wu, Zhi Zheng, Zhaopeng Qiu, Hao Wang, Hongchao Gu, Tingjia Shen, Chuan Qin, Chen Zhu, Hengshu Zhu, Qi Liu, et al. A survey on large language models for recommendation. volume 27, article number 60. Springer, 2024.
- [7] Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. Graph neural networks in recommender systems: A survey. volume 55, pages 1–37. ACM, 2022.
- [8] Zheyuan Liu, Xiaoxin He, Yijun Tian, and Nitesh V. Chawla. Can we soft prompt LLMs for graph learning tasks? In *Proceedings of the ACM Web Conference 2024 (WWW)*, pages 481–484, 2024.
- [9] Brian Lester, Rami Al-Rfou, and Noah Constant. The power of scale for parameter-efficient prompt tuning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3045–3059, 2021.
- [10] Michael J Pazzani and Daniel Billsus. Content-based recommendation systems. In *The adaptive web: methods and strategies of web personalization*, pages 325–341. Springer, 2007.
- [11] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3982–3992, 2019.
- [12] Peipei Xia, Li Zhang, and Fanzhang Li. Learning similarity with cosine similarity ensemble. volume 307, pages 39–52. Elsevier, 2015.
- [13] Y. Ding, B. Wang, X. Cui, and M. Xu. Popularity prediction with semantic retrieval for news recommendation. volume 247, article number 123308, 2024.
- [14] Huifeng Guo, Ruiming Tang, Yunming Forest Ye, Zhenguo Li, and Xiuqiang He. DeepFM: A factorization-machine based neural network for CTR prediction. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1725–1731, 2017.
- [15] Zeno Gantner, Lucas Drumond, Christoph Freudenthaler,

- Steffen Rendle, and Lars Schmidt-Thieme. Learning attribute-to-feature mappings for cold-start recommendations. In *Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM)*, pages 176–185. IEEE, 2010.
- [16] Hoyeop Lee, Jinbae Im, Seongwon Jang, Hyunsouk Cho, and Sehee Chung. MeLU: Meta-learned user preference estimator for cold-start recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1073–1082, 2019.
- [17] Yinwei Wei, Xiang Wang, Qi Li, Liqiang Nie, Yan Li, Xuanping Li, and Tat-Seng Chua. Contrastive learning for cold-start recommendation. In *Proceedings of the 29th ACM international conference on multimedia*, pages 5382–5390, 2021.
- [18] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. Neural graph collaborative filtering. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pages 165–174, 2019.
- [19] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*, 2017.
- [20] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*, volume 30, pages 1025–1035, 2017.
- [21] Jinri Kim, Eungi Kim, Kwangeun Yeo, Yujin Jeon, Chanwoo Kim, Sewon Lee, and Joonseok Lee. Content-based graph reconstruction for cold-start item recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pages 1263–1273, 2024.
- [22] Wei Jin, Yao Ma, Xiaorui Liu, Xianfeng Tang, Suhang Wang, and Jiliang Tang. Graph structure learning for robust graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, pages 66–74, 2020.
- [23] Zihuai Zhao, Wenqi Fan, Jiatong Li, Yunqing Liu, Xiaowei Mei, Yiqi Wang, Zhen Wen, Fei Wang, Xiangyu Zhao, Jiliang Tang, et al. Recommender systems in the era of large language models (LLMs). volume 36, pages 6889–6907. IEEE, 2024.
- [24] Peiyuan Zhang, Guangtao Zeng, Tianduo Wang, and Wei Lu. Tynllama: An open-source small language model. *arXiv preprint arXiv:2401.02385*, 2024.
- [25] Yupeng Hou, Junjie Zhang, Zihan Lin, Hongyu Lu, Ruobing Xie, Julian McAuley, and Wayne Xin Zhao. Large language models are zero-shot rankers for recommender systems. In *European Conference on Information Retrieval*, pages 364–381. Springer, 2024.
- [26] Keqin Bao, Jizhi Zhang, Yang Zhang, Wenjie Wang, Fuli Feng, and Xiangnan He. TALLRec: An effective and efficient tuning framework to align large language model with recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems (RecSys)*, pages 1007–1014, 2023.
- [27] Yangqin Jiang, Yuhao Yang, Lianghao Xia, Da Luo, Kangyi Lin, and Chao Huang. Reclm: Recommendation instruction tuning. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15443–15459, 2025.
- [28] Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*, pages 4582–4597, 2021.
- [29] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 452–461, 2009.
- [30] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. volume 55, pages 1–35. ACM New York, NY, 2023.
- [31] Maksims Volkovs, Guangwei Yu, and Tomi Poutanen. DropoutNet: Addressing cold start in recommender systems. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*, volume 30, pages 4964–4973, 2017.
- [32] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2015.

適応的回転ゲートと局所探索を用いた 量子ビット進化アルゴリズムによる TGNN の説明手法

小原 陸哉[†] 小島 豊[†] 佐々木勇和[†]

[†] 大阪大学大学院情報科学研究科 〒 565-0871 大阪府吹田市山田丘 1-5

E-mail: †{ohara.rikuya,kojima.yutaka,sasaki}@ist.osaka-u.ac.jp

あらまし 時間グラフニューラルネットワーク (TGNN) は動的グラフにおけるイベントの予測に優れる一方、その判断根拠が不透明であるという課題がある。そこで、モデルの予測の根拠となる部分グラフを特定して提示する説明手法が求められる。TGNN の説明生成手法である既存手法 QIEA-TGX は、量子ビット表現に基づく進化アルゴリズムを用いることで、広大な探索空間においても多様性を維持しつつ効率的に探索が可能である。この手法のさらなる精度向上を図るため、適応的回転行列と局所探索による新たな手法 QIEA-TGX+ を提案する。実験では四つの実データセットを用いて、六つのベースライン手法と比較を行い、結果として、提案手法は既存手法と比較し、事実説明では約 0.81%、反実仮想説明では約 1.14%の説明精度の向上を達成したことを示す。

キーワード TGNN, 説明手法, QIEA-TGX

1 はじめに

時系列ネットワーク (Temporal Network) とは、構造と特性が時間の経過とともに変化するネットワークである [1]。システムの構成要素を表すノード集合と、ノード間の相互作用を表すリンク集合という二つの主要素で構成される。時系列ネットワークでは、時間の経過とともに、新たなノードやノード間のリンクの発生が起り、ある時刻にリンクが発生する事象を「イベント」として表現する。こうした動的特性に基づき、ネットワーク内の将来のリンクを予測する「リンク予測」は、ソーシャルネットワーク [2] や通信ネットワーク [3]、交通予測 [4,5]、知識グラフ [6,7] といった様々な分野への応用が可能である。しかし、従来のグラフニューラルネットワーク (GNN) が扱うことのできるグラフは静的グラフに限定されており、動的グラフを扱うことのできる新たな学習モデルが必要となった。

その問題を解決するため、時間グラフニューラルネットワーク (TGNN) が提案された [8]。TGNN は、GNN を拡張し、時間的な依存関係を学習することで動的なノード、エッジ表現の学習を可能にした深層学習モデルであり、システム内の空間的な依存関係と時間的な変化の両方を捉えることが可能である [8]。TGNN は時間グラフにおけるリンク予測やノード分類といったタスクにおいて高い性能を発揮する。そして、その予測や分類の根拠を提示する説明可能性は、TGNN の重要な要素である。これは、TGNN の説明可能性は TGNN モデルを実世界に展開する際の信頼性に関わり、また、モデルの透明性が高いとシステム内の動的な依存関係を識別しやすくなるためである。したがって、これまで様々な TGNN の説明手法が提案されてきた [9-11]。

QIEA-TGX は、量子ビット表現に基づく進化アルゴリズムである Quantum-Inspired Evolutionary Algorithm (QIEA) [12] を用いた TGNN の説明手法である [13]。TGNN における説

明手法は、時間グラフの複雑性のために計算コストが高くなりやすいが、QIEA-TGX はこの問題に対する効率的な解法を提供する。QIEA-TGX は量子ビットによって個体表現を行うため、一度に複数の解を表現可能であり、同時に局所最適解を減らすことが可能である。作成された説明グラフを TGNN に入力して評価値を取得し、その評価値に基づいて量子ゲートによる量子ビットの更新を繰り返し、説明ごとの最良の解を探索する。また、探索時にはターゲットイベントの 1hop 近傍を先に評価し、不要な 2hop 近傍を削減して評価するため、不要な探索を削減し、近傍イベントのより適切な評価を可能とする。この大域的かつ効率的な探索能力により、QIEA-TGX は、実世界のデータセットを用いた実験において、最先端のベースラインと比較し、説明の忠実度の向上を達成した。

しかし、QIEA-TGX には主に二つの課題が存在する。一つ目は、量子ビットの更新時に用いる回転ゲートの回転角の大きさが固定されている点である。更新幅が毎回同じ大きさだと、グラフに応じた最適更新ができず、最適解への収束が難しい。二つ目は、QIEA の大域的な探索後でも、存在確率が「0」と「1」の間で揺れ動き、必要性がうまく確定しなかったイベントが存在することがある点である。このように QIEA では決定できなかったイベントに対しては、別の機構でその必要性を決定させる必要がある。

本研究では、QIEA-TGX のさらなる説明精度向上を目的として、QIEA-TGX+ を提案する。QIEA-TGX+ には、前述した二つの課題を解決するための二つの改良が含まれる。一つ目は、適応的な探索を実現する Adaptive Rotation Gate Operator (ARGO) 機構の導入である。ARGO は、QIEA では固定されていた回転ゲートの回転角を、その時点の個体群の評価値に基づいて動的に調整する機構である [14]。新しく得られた解が最適解に近い場合、回転角は細分化され、遠い場合、探索空間をより深く探索するため、回転角を大きくする。現在の解と最適解に基づいて回転角を調整するため、説明対象のイベントである

ターゲットイベントに対して、より適した収束を促進できる。また、これらの角度は最小値と最大値の間で固定されるため、アルゴリズムの発散や早期収束も防ぐことができる。

二つ目は、QIEA の探索後に得られた暫定最良解に対し、局所探索を行う二段階の探索機構である。QIEA の大域的な探索後でも必要性がうまく確定しなかったイベントに対し、各イベントの存在確率を反転させ、説明精度が上昇するような組み合わせを探索する。QIEA の探索後に局所探索を行うような、二段階の探索機構に設計することで、QIEA では必要性の判定が困難だったイベントの選別精度が向上し、全体的な説明精度の向上が期待できる。

実験では、四つの実データセット (Wikipedia, Reddit, MovieLens, Enron) を用いて、提案手法 QIEA-TGX+ と、六つのベースライン手法 (random, T-GNNExplainer [9], GreeDy [11], PGExplainer [15], GA-TGX [13], QIEA-TGX [13]) との比較を行う。評価方法は、存在するイベント、存在しないイベントに対し、事実説明および反実仮想説明を個別に実施し、説明の寄与度 (fidelity)、実行時間を計算する。結果として、既存手法と比較し、事実説明では約 0.81%、反実仮想説明では約 1.14% の説明精度の向上を達成したことを示す。

2 関連研究

2.1 時間グラフニューラルネットワーク

時間グラフは、離散時間動的グラフ (DTDG) と連続時間動的グラフ (CTDG) の二つに分けることができる [1]。DTDG は、時間グラフをタイムスタンプ付きのスナップショット集合に簡略化し、連続的な時空間を離散的なタイムスタンプに変換する。CTDG は、時空間を離散化せず、グラフをタイムスタンプ付きのイベントのリストで表し、時系列ネットワークの時間的連続性を維持する。本研究における時間グラフは CTDG を対象とする。

CTDG モデルでは、イベントや履歴ノード情報を近傍のノードから集約してノード埋め込みを更新する手法が主流である。代表的な手法として、自己注意機構により時間情報を集約する TGAT [16] や、履歴情報を保持するメモリモジュールを導入した TGN [17] が挙げられる。近年の研究では、これらの手法の効率化や大規模言語モデル (LLM) への応用といった試みが行われている。DG-Mamba [18] は、状態空間モデル (Mamba) を導入することで TGNN の課題である計算コストの削減を図った CTDG モデルである。計算負荷の高い Transformer の代わりに Mamba を用いることで、時間計算量を二次から線形へと削減できる。LLM4DyG [19] は、LLM と動的グラフを組み合わせたベンチマークであり、LLM の推論能力を使って動的グラフの時空間パターンを解釈させる仕組みを提案している。

2.2 グラフニューラルネットワークの説明手法

GNN の説明の種類は、事実説明と反実仮想説明の大きく二つに分類できる [20]。事実説明は、予測に最大の影響を与えるノード特徴量または部分グラフを見つけることを目指す。一方、

反実仮想説明は、モデルの予測を覆す、入力グラフへの最小の変更を見つけることで説明を提供する。

事実説明の説明手法である GNNExplainer [21] は、入力の一部を変化させ予測出力を観察することで、予測に最大の影響を与える部分グラフを見つける。反実仮想説明の説明手法である CF-GNNExplainer [22] は、摂動後の反実仮想グラフの予測精度と、元のグラフとの距離を考慮した損失関数を最適化することで説明を生成する。また、動的グラフに対して用いられる説明手法として、T-GNNExplainer [9] が挙げられる。T-GNNExplainer は静的グラフ用の GNNExplainer を TGNN に拡張した事実説明手法であり、モンテカルロ木探索を用いて重要なイベント集合を探索する。また、近年提案された TempME [10] は予測に最も寄与する頻出パターンを抽出することで、より解釈性の高い説明を提供する。さらに、近年の研究では、生成 AI 技術の応用が進んでいる。Graph Diffusion Counterfactual Explanation [23] は、拡散モデル (Diffusion Model) を用いてグラフのデータ分布を学習し、より自然な反実仮想グラフを生成する。LLM-GCE [24] は、グラフ説明手法と LLM を統合し、予測の根拠やグラフの構造変化を自然言語で説明させる試みを提案している。このように、最近の生成 AI 技術の進歩により、説明手法も新たな進展を見せている。

2.3 Quantum-Inspired Evolutionary Algorithm

Quantum-Inspired Evolutionary Algorithm (QIEA) は、量子ビットや状態の重ね合わせといった量子コンピューティングの概念と原理に基づいた進化アルゴリズムである [25]。進化アルゴリズムにおける各個体を量子ビットによる重ね合わせ状態で表現することで、広大な探索空間においても多様性を維持し、効率的に探索が可能である。

QIEA アルゴリズムには、様々な改良が行われている。ARGO [14] は、心音信号から雑音を除去するフィルタの係数を最適化するために QIEA に導入された新たな手法であり、解の評価値に基づいて、量子ゲートの回転角を動的に調整する。最適解に近いときは細かく探索し、遠いときは広く探索するように角度を変えることで、最適な収束が可能になる。QIOA [26] は、QIEA にトンネル効果や量子もつれなどの量子力学的現象を組み込んだ手法である。トンネル効果に基づき、「エネルギー障壁」を透過して向こう側へ移動する確率関数を導入し、局所解からの脱出能力を強化したり、量子もつれを通じて、ある個体が良い解を見つけると、他の個体もその情報を踏まえた動きに調整され、効率的な探索を可能になるなど、QIEA の性能向上に成功している。

また、最新の研究では、その汎用性から QIEA は様々な場面に適用されている。CNN モデルのハイパーパラメータチューニングでは、QIEA は従来の遺伝的アルゴリズムやベイズ最適化を上回る精度と収束速度を達成した [27]。ビッグデータ解析においても、初期段階で多様な組み合わせを探索し徐々に有望なサブセットに収束する挙動が極めて相性がいいとされ、QIEA を用いた手法が、様々な高次元データセットにおいて特徴数を大幅に削減しつつ、分類精度を向上させた [28]。

3 事前知識

3.1 時間グラフと時間グラフニューラルネットワーク

連続時間動的グラフ (CTDG) は, $G = (\mathcal{N}, \mathcal{E})$ と表せる. ここで, \mathcal{N} と $\mathcal{E} = \{e_1, e_2, \dots, e_T\}$ はそれぞれノードとイベントの集合を表す. 各イベント $e_i = (u_i, v_i, t_i)$ は, タイムスタンプ t_i に発生する, ソースノード u_i からターゲットノード v_i へのインタラクションを表す. 時間グラフニューラルネットワーク (TGNN) におけるリンク予測は, グラフ $G^k = (\mathcal{N}^k, \mathcal{E}^k)$ ($\mathcal{E} = \{e_1, e_2, \dots, e_{k-1}\}$) から, 与えられたリンク e_k の発生を予測することを目的としている. 時刻 t_k までの時間グラフ G^k 内の全イベント集合 \mathcal{E}^k から, ターゲット e_k の近傍に存在するイベント集合 \mathcal{E}^c ($\mathcal{E}^c \subset \mathcal{E}^k$) を抽出し, ターゲット e_k と G^c を学習済みモデル f に入力することで, 予測確率 $Y_k = f(G^c)[e_k]$ を得る.

3.2 問題定義

TGNN のリンク予測における説明では, 予測の根拠となる重要なイベントを含む説明グラフ $G^e = (\mathcal{N}^e, \mathcal{E}^e)$ を特定することを目的としている. 説明には事実説明と反実仮想説明があり, 対象イベントは存在する場合と存在しない場合がある. したがって, 説明には四つの種類がある. 時系列グラフ $G^c \subseteq G^k$, 対象イベント e_k , TGNN モデル f , sparsity $\phi \in (0, 1)$ が与えられた場合, 以下の G^e を特定することを目的とする.

- 存在するイベントの事実説明:

$$\text{maximizes } f(G^e)[e_k] \text{ subject to } \frac{|\mathcal{E}^e|}{|\mathcal{E}^c|} \leq \phi$$
 - 存在しないイベントの事実説明:

$$\text{minimizes } f(G^e)[e_k] \text{ subject to } \frac{|\mathcal{E}^e|}{|\mathcal{E}^c|} \leq \phi$$
 - 存在するイベントの反実仮想説明:

$$\text{minimizes } f(G^c \setminus G^e)[e_k] \text{ subject to } \frac{|\mathcal{E}^c \setminus \mathcal{E}^e|}{|\mathcal{E}^c|} \leq \phi$$
 - 存在しないイベントの反実仮想説明:

$$\text{maximizes } f(G^c \setminus G^e)[e_k] \text{ subject to } \frac{|\mathcal{E}^c \setminus \mathcal{E}^e|}{|\mathcal{E}^c|} \leq \phi$$
- ここで, $G^c \setminus G^e = (\mathcal{N}^c, \mathcal{E}^c \setminus \mathcal{E}^e)$ である. sparsity ϕ は説明グラフに使用するイベント数を制限するために設定し, 元グラフのイベント集合 \mathcal{E}^c 内のイベントと ϕ の積の値が選択可能な最大イベント数である.

事実説明は, 元の予測結果を再現する最小限の部分グラフを説明グラフ G^e として提示する. 存在するイベントに対する説明の場合, 予測値を最も高く出力可能な説明グラフ G^e を探索する. 存在しないイベントに対する説明の場合, 予測値を最も低く出力可能な説明グラフ G^e を探索する.

反実仮想説明は, 入力に対する最小限のイベント削除によって, 予測が覆るような部分グラフを説明グラフ G^e として提示する. 存在するイベントに対する説明の場合, イベントの削除によって予測値を最も低く出力する説明グラフ G^e を探索する. 存在しないイベントに対する説明の場合, イベントの削除によって予測値を最も高く出力する説明グラフ G^e を探索する.

3.3 QIEA-TGX

QIEA-TGX [13] は, TGNN のリンク予測結果に対する説明

生成手法である. Quantum-Inspired Evolutionary Algorithm (QIEA) を応用して, イベントの組み合わせを効率的に探索する. 図 1 に QIEA-TGX の概略図を示す. 図 1 における QIEA-TGX の全体的な流れは以下の通りである.

- 初期化:** 全てのイベントの存在確率 (量子ビット) を均等な確率 (0.5) で初期化.
- 観測:** 量子ビットを観測し, バイナリビットに変換して存在, 不存在を確定.
- 評価:** TGNN にグラフ (バイナリビット列) を入力し, 適応度を計算.
- 更新:** 評価結果に基づき, ゲート操作で量子ビットを更新.
- 出力:** 終了条件を満たすまで 2, 3, 4 の手順を繰り返し, 適応度の最も高いグラフを説明グラフとして出力.

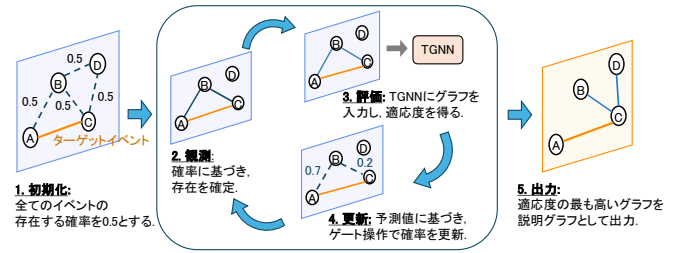


図 1: QIEA-TGX の概略図

まずは, QIEA-TGX における個体表現について説明する. 従来の進化アルゴリズムでは, 個体は通常, 2 進数, 実数, または記号的な表現によって符号化される [29]. 一方, QIEA では確率的な状態を持つ量子ビットで個体を表現する. 量子ビットは次のように表現する.

$$|q\rangle = \alpha|0\rangle + \beta|1\rangle$$

ここで, α と β は確率振幅であり, 正規化条件 $|\alpha|^2 + |\beta|^2 = 1$ を満たす. $|\alpha|^2$ および $|\beta|^2$ は, それぞれ量子ビットが「0」, 「1」として観測される確率を表す. 長さ n の量子ビットにおける個体 Q は, 以下のように表現される.

$$Q = \{|q_1\rangle, |q_2\rangle, \dots, |q_n\rangle\}$$

この表現により, 各量子ビットは「0」および「1」の両方の状態を同時に保持でき, 2^n 通りの状態を同時に表現することが可能となる. このように, 一つの個体が複数の解の状態を同時に表現することで, 探索の多様性と効率性が大幅に向上する.

次に, QIEA-TGX の操作手順について説明する. まずは, 各量子ビットを「0」と「1」が等確率で存在する重ね合わせ状態に初期化する. 具体的には, 長さ n の量子ビット列において, 各量子ビット $|q_i\rangle$ を以下の式で初期化する.

$$|q_i\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle \quad (i = 1, 2, \dots, n)$$

ここで係数 $\frac{1}{\sqrt{2}}$ は, 観測確率 $|\alpha|^2 = |\beta|^2 = 0.5$ を意味する. これにより, 探索空間全体を均等に探索することが可能となり, 初期段階での解空間の多様性が確保される. 本研究では, 量子

ビット列の長さ n はイベント集合内のイベント数 $|\mathcal{E}^c|$ に対応する。

次に、観測操作を行う。観測では、量子ビットの状態を具体的なバイナリビットに変換する。各量子ビット $|q_i\rangle = \alpha_i|0\rangle + \beta_i|1\rangle$ は、以下の確率に基づいて「0」または「1」に決定される。

$$P(0) = |\alpha_i|^2, \quad P(1) = |\beta_i|^2$$

この観測によって生成されたバイナリビット列が評価対象のグラフとなり、適応度の評価に使用される。

次に、評価操作を行う。観測によって確定したバイナリビット列を TGNN に入力し、その出力を適応度とする。適応度とは、生成されたグラフがターゲットイベントの予測にどれだけ寄与しているかを示す指標であり、具体的には TGNN によるリンク予測の予測値等に対応する。

次に、更新操作を行う。QIEA-TGX における量子ビットの更新操作は、Q ゲートによる更新と H ゲートによる更新の二つが存在する。

Q ゲートによる更新は、回転操作を通じて量子ビットの確率振幅 α, β を調整し、解の確率分布を動的に変化させる操作である。新たに観測された個体の適応度が、その時点での観測された最良個体の適応度よりも高く、かつ対応するビットが異なる場合、量子ビットの状態を最良個体に近づけるように回転させる。回転行列 $R(\theta_Q)$ は以下のように定義される。

$$R(\theta_Q) = \begin{bmatrix} \cos(\theta_Q) & -\sin(\theta_Q) \\ \sin(\theta_Q) & \cos(\theta_Q) \end{bmatrix}$$

これを用いて、量子ビットは以下のように更新される

$$\begin{bmatrix} \alpha' \\ \beta' \end{bmatrix} = R(\theta_Q) \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

回転行列を用いることで、更新された量子ビット $|q'_i\rangle = \alpha'_i|0\rangle + \beta'_i|1\rangle$ は正規化条件を満たす。この操作により、優良な解の出現確率を増加させ、大域的な最適解の探索が促進される。

H ゲート (Hadamard ゲート) による更新は、量子ビットの基底状態を重ね合わせ状態に変換し、探索空間の多様性を向上させる重要な役割を担う [30]。H ゲートは、量子ビットを操作して「0」と「1」の両方の状態が同時に存在する重ね合わせの状態を生成するため、「0」か「1」の両者が観測される最低確率を保ち、探索の多様性を確保できる。QIEA-TGX では、回転操作により確率振幅が極端に偏った場合に、確率振幅を強制的に $\sqrt{\theta_H}$ に設定する。例えば、 α^2 が 0 に近づいた場合、つまり特定の閾値 $\sqrt{\theta_H^2}$ を下回った場合、H ゲートを適用して、 α^2 を $\sqrt{\theta_H^2}$ に、 β^2 を $\sqrt{1-\theta_H^2}$ に設定する。このように H ゲートの適用に条件を付けることで、各量子ビットの偏りを緩和し、特定の基底状態への収束を防ぐことができる。

以上の「観測」、「評価」、「更新」からなる一連の探索プロセスを一定回数だけ繰り返す。その過程で、観測時にそれまでに最も高い適応度を示した個体が、暫定的な最良解として各回(世代)で更新され続ける。一定回数の反復が終了した時点で、

観測結果が最も高い適応度となる個体をバイナリビット列に変換し、説明グラフとして出力する。この説明グラフは、予測モデルの判断に最も寄与したイベント集合であり、予測の根拠として提示する。

4 提案手法

本研究では、QIEA-TGX に対し「ARGO による適応的更新」、「局所探索による事後最適化」の二つの改良機構を導入した手法 QIEA-TGX+ を提案する。図 2 に QIEA-TGX+ の概略図を示す。

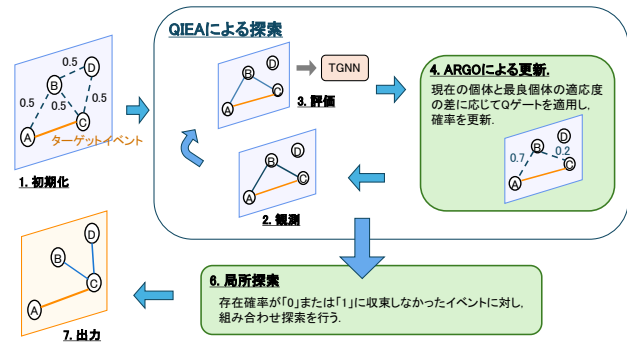


図 2: QIEA-TGX+の概略図

具体的な処理の流れは以下の通りである。まず、説明対象となるグラフに含まれるイベントの存在確率を 0.5 で初期化する。次に、QIEA-TGX の探索プロセスにおいて、各個体の観測、TGNN による評価、そして ARGO を用いた量子ゲート更新の三つの操作を繰り返す。一定回繰り返された時点で終了し、最後に、QIEA-TGX による探索で得られた暫定解に対し、確率が収束しなかったイベントのみを対象とした局所探索を行い、最終的な説明グラフを出力する。各改良機構の詳細は以降の節で示す。

4.1 ARGO

Adaptive Rotation Gate Operator (ARGO) [14] は、QIEA の量子ビットの回転操作における改良手法である。従来の QIEA における回転ゲートは一般に固定された回転角を用いるが、ARGO では個体の適合度に応じて回転角を変化させる。この設計は、大域的な探索と局所的な深化のバランスを動的に調整し、解の早期収束を可能にする。また、回転角がある一定の範囲に収まる設計により、アルゴリズムの安定性が確保される。ARGO における回転角 θ_Q の計算は以下のように表される。

$$\theta_Q = \begin{cases} \theta_Q \times s, & \text{when bits differ} \\ \theta_Q, & \text{when bits do not differ} \end{cases}$$

$$\theta_Q = \theta_{\min} + \left| \frac{f_{\text{best}} - f_i}{K} \right| \times (\theta_{\max} - \theta_{\min})$$

ここで、 θ_{\min} および θ_{\max} はそれぞれ回転角の下限値と上限値であり、回転角はこの範囲内で制御される。 K は適応度の差分を正規化するためのスケール係数である。適応度の差を

アルゴリズム 1 ARGO による個体の更新手順

入力: 候補イベント集合 S^c , 個体 Q , 個体のバイナリビット列 P , 最良のバイナリビット列 B , 各個体の適応度 f , 最良適応度 f_{best} , 回転角の最小値 θ_{min} , 回転角の最大値 θ_{max} , 正規化係数 K

出力: 更新された個体 Q

```

1:  $\Delta f \leftarrow |f_{best} - f|$ 
2:  $\theta_{argo} \leftarrow \theta_{min} + (\Delta f / K) \times (\theta_{max} - \theta_{min})$    ▷ 回転角を計算
3: if  $f < f_{best}$  then   ▷ 各個体の最良適応度より悪い場合のみ更新
4:   for  $j \leftarrow 1$  to  $|S^c|$  do
5:     if  $P[j] \neq B[j]$  then   ▷ 観測ビットと最良ビットが異なる場合
6:       if  $P[j] = 0$  and  $B[j] = 1$  then
7:          $s \leftarrow 1$    ▷ 正方向へ回転
8:       else if  $P[j] = 1$  and  $B[j] = 0$  then
9:          $s \leftarrow -1$    ▷ 負方向へ回転
10:      end if
11:       $Q[j] \leftarrow Q\_gate(Q[j], s \times \theta_{argo})$    ▷ 最良個体に近づくように回転
12:    end if
13:  end for
14: end if
15:  $Q \leftarrow H\_gate(Q)$    ▷ H ゲートの適用
16: return  $Q$ 

```

取ることにより、最良個体から遠い個体には大きな回転角が適用され、大域的な「探索」が促進される。一方、最良個体に近い個体には小さな回転角が適用され、解の近傍を細かく探索する「深化」が行われる。 s は量子ビットが異なる場合の回転方向を決定するための符号で、現在の個体の量子ビットの観測値 $Obs(|q_{new}\rangle)$ と最良個体の量子ビット $Obs(|q_{best}\rangle)$ の値に基づいて、以下のように決定される。

$$s = \begin{cases} 1, & Obs(|q_{new}\rangle) = 0, Obs(|q_{best}\rangle) = 1 \\ -1, & Obs(|q_{new}\rangle) = 1, Obs(|q_{best}\rangle) = 0 \end{cases}$$

すなわち、 s により、現在の観測値が最良解と一致している場合は回転を行わず、異なる場合は最良解へ近づく方向へ回転を加える。この適応的な制御により、探索の初期段階では多様性を維持しつつ、終盤では高精度な収束を可能にする。

ARGO を導入した QIEA の更新機構のアルゴリズムをアルゴリズム 1 に示す。入力の P は個体を観測してバイナリビットに変換したビット列、 B は最良個体を観測してバイナリビットに変換したビット列である。まず、2 行目と 3 行目で ARGO における回転角 θ_Q を計算する。4 行目と 6 行目では Q ゲートを適用する際の条件判定を行っている。更新は適応度が各個体の最良適応度より悪い個体のみに行う。また、観測したビットと最良個体のビットを比較して、異なる量子ビットのみに更新を行う。7 行目から 11 行目では s を計算し、ビットが異なる場合の回転方向を決定する。最後に、12 行目で Q ゲート、16 行目で H ゲートを適用し、更新操作を終了する。

4.2 局所探索

QIEA による大域的な探索が終了した後、すべての量子ビッ

トが 0 または 1 に完全に収束するとは限らず、0.5 付近で停滞し必要性が定まっていないビットが存在する場合がある。このようなビットを確率的に観測して最終解とすることは、最終的な説明グラフの評価を低下させるリスクがある。

この問題に対処するため、QIEA の探索後も必要性が確定しなかったイベントに対し、局所的な最適化を行い、その必要性を確定させる。まず、QIEA による探索で得られた最良個体 Q_{best} の量子ビットから存在確率 β^2 を算出し、未確定イベント e の集合 U を特定する。

$$U = \{e_j \in Q_{best} \mid \tau_{low} < \beta_j^2 < \tau_{high}\}$$

ここで τ_{low}, τ_{high} は閾値であり、存在確率がこの範囲内に収まるイベントを未確定として扱う。その後、 U に含まれる未確定イベント e_j について、ビットを反転させたときの適応度の変化量 Δf を計算し、 e_j が 0 なら反転ビット 1 を追加候補リスト L_{add} に、 e_j が 1 なら反転ビット 0 を削除候補リスト L_{drop} に Δf と一緒に追加する。すなわち、 L_{add} は、これを選ぶとイベントの使用数が増加 ($0 \rightarrow 1$) する e_j 、 L_{drop} は、これを選ぶとイベントの使用数が減少 ($1 \rightarrow 0$) する e_j である。二つのリストは Δf の降順にソートを行う。リスト作成後は、以下の二つの段階で処理を行う。

- 追加段階:** 現在のイベントの使用数がイベントの最大採用可能数 \max_{event} 以下なら、イベント数が \max_{event} になるまで、 $\Delta f > 0$ である L_{add} 内のイベントを上位から順にグラフに追加する。
- 入れ替え段階:** L_{add} 内のイベントを採用したときの適応度上昇幅 Δf_{inc} と L_{drop} 内のイベントを採用したときの適応度減少幅 Δf_{dec} の和を計算し、 $\Delta f_{inc} + \Delta f_{dec} > 0$ となる組み合わせに対し、 L_{add} 内のイベントを追加、 L_{drop} 内のイベントを削除し、入れ替えを行う。

すなわち、未確定イベント e_j を反転し、適応度が上昇するなら、 \max_{event} の範囲内でその反転を採用する。また、反転して適応度が上昇するイベントと減少するイベントが二つ存在するなら、それらのイベントを入れ替える。この事後的な探索により、QIEA が大域的に絞り込んだ探索空間の中で、最終的な説明サブグラフを修正し、局所最適解への到達を確実にする。

局所探索機構のアルゴリズムをアルゴリズム 2 に示す。まず、1 行目で最良個体 Q_{best} の量子ビットから存在確率 β^2 を算出し、2 行目で未確定イベント e の集合 U を特定する。その後、 U に含まれる未確定イベント e_j についてのみ、5 行目以降で順次値を確定させる。まず、7 行目で、未確定イベント e_j を反転した場合の適応度を計算する。そして、10 行目、12 行目で、 e_j が 0 なら追加候補、1 なら削除候補のリストに追加する。これを U に含まれるすべての e_j に対して行う。15 行目ではリストのソートを行う。16 から 24 行目は、イベントの追加段階である。現在の最良個体を観測し、採用イベント数が最大値以下で、かつ採用により適応度が上昇するイベントが追加候補リストに存在するなら、20 行目でそのイベントを採用し、21 行目で適応度を更新する。イベント追加後、25 から 36 行目はイベントの入れ替え段階である。追加候補リストのイベント追加による適応

アルゴリズム 2 局所探索の手順

入力: 個体 Q , 最良のバイナリビット列 B , 最良のインデックス $best$, 最良適応度 f_{best} , 閾値 τ_{low}, τ_{high} , 最大イベント数 max_{event}

出力: 修正後の最良のバイナリビット列 B

```

1:  $P \leftarrow \sin^2(Q[best])$  ▷ 確率計算
2:  $U \leftarrow \{j \mid \tau_{low} < P[j] < \tau_{high}\}$  ▷ 未確定ビット集合
3:  $L_{add} \leftarrow \emptyset, L_{drop} \leftarrow \emptyset$ 
4: for  $j \in U$  do
5:    $B' \leftarrow B$ 
6:    $B'[j] \leftarrow 1 - B[j]$  ▷ ビット反転
7:    $\Delta f \leftarrow \text{Evaluate}(B') - f_{best}$  ▷ 反転した場合の適応度を計算
8:   // 最小化問題なら逆の操作を行う
9:   if  $B[j] = 0$  then
10:     Append  $(j, \Delta f)$  to  $L_{add}$  ▷ 0→1: 追加候補リストへ
11:   else
12:     Append  $(j, \Delta f)$  to  $L_{drop}$  ▷ 1→0: 削除候補リストへ
13:   end if
14: end for
15: Sort  $L_{add}$  and  $L_{drop}$  by  $\Delta f$  in descending order
16: // 追加段階
17: for  $(j, \Delta f) \in L_{add}$  do
18:    $k \leftarrow \text{CountOnes}(B)$ 
19:   if  $k < max_{event}$  and  $\Delta f > 0$  then
20:      $B[j] \leftarrow 1$  ▷ イベントを追加
21:      $f_{best} \leftarrow f_{best} + \Delta f$  ▷ 適応度を更新
22:     Remove  $(j, \Delta f)$  from  $L_{add}$ 
23:   end if
24: end for
25: // 入れ替え段階
26: for  $(i, \Delta f_{inc}) \in L_{add}$  do
27:   for  $(j, \Delta f_{dec}) \in L_{drop}$  do
28:     if  $\Delta f_{inc} + \Delta f_{dec} > 0$  then ▷ 適応度が上がる場合
29:        $B[i] \leftarrow 1 - B[i]$ 
30:        $B[j] \leftarrow 1 - B[j]$  ▷ イベントを入れ替え
31:        $f_{best} \leftarrow f_{best} + (\Delta f_{inc} + \Delta f_{dec})$  ▷ 適応度を更新
32:       Remove  $(j, \Delta f_{dec})$  from  $L_{drop}$ 
33:       break ▷ ペア成立のため次の追加候補へ
34:     end if
35:   end for
36: end for
37: return  $B$ 

```

度上昇量と削除候補リストのイベント削除による適応度減少量の和が正である場合, 29, 30 行目でイベントの入れ替えを行い, 31 行目で適応度を更新する. これを適応度の変化幅が正である限り繰り返し, すべてのイベントの判定を終えた時点での個体を, 最終的な説明サブグラフとする.

5 実験

本章では, [13] で用いられた六つのベースライン手法に加え, 提案手法である QIEA-TGX+ の計七つの手法で精度の検証を行う. 実験は, Intel Xeon Silver 4210R CPU(40 コア, 2.40 GHz), NVIDIA RTX 6000 GPU (40 GB) を搭載したサー

表 1: データセットの概要

データセット	ソース	ターゲット	合計ノード	リンク	特徴量の次元
Wikipedia	8,227	1,000	9,227	157,474	172
Reddit	10,000	984	10,984	672,447	172
MovieLens	200,948	87,585	288,533	32,000,204	1
Enron	158	158	158	517,431	172

バ (OS: Ubuntu 22.04.4 LTS) で行った.

5.1 実験設定

5.1.1 時間グラフニューラルネットワークとデータセット
 本研究では, 説明対象の TGNN として TGAT [16] と TGN [17] を用いた. データセットには, Wikipedia, Reddit, Movielens, Enron を用いた. データセットの概要は表 1 に示す. TGNN の学習は, 全データを時系列順に 70% : 15% : 15% の割合で分割し, それぞれを訓練データ, 検証データ, テストデータとして用いた.

5.1.2 比較手法

提案手法 QIEA-TGX+ に加えて, 六つの比較手法で実験を行う. 比較手法を以下に示す.

- **random:** sparsity の制約を満たすようにランダムにイベントをサンプリングし, 評価を規定回数繰り返して, 最適なイベント集合を作成する.
- **T-GNNEExplainer [9]:** PGExplainer [15] をナビゲータとして利用し, モンテカルロ木探索によって最適なイベント集合を探索する.
- **GreeDy [11]:** イベントを 1 本ずつ追加, 削除した場合の推論結果を確認し, 予測値が最も良くなるイベントの追加, 削除を繰り返す.
- **PGExplainer [15]:** 入力グラフの各イベントに対する重要度を推定し, 重要度が高い順にイベントを説明グラフに追加する.
- **GA-TGX [13]:** 遺伝アルゴリズムをベースに, 交叉や突然変異操作を通して最適なイベント組み合わせを探索する.
- **QIEA-TGX [13]:** QIEA による量子ビット表現と量子ゲート操作により, 多様な解表現で効率的に最適なイベント組み合わせを探索する.

比較手法と実験設定は [13] と同様である. 全ての説明手法の最大実行時間を 10 分とし, sparsity は 0.2 とした. random, GA-TGX, QIEA-TGX は説明グラフの作成回数は 500 回に限定する. GreeDy は, イベントの追加または削除によって予測値が改善しなくなるか, sparsity によるイベント数上限に達した場合に実行を終了する. さらに, TGNN の学習には 2hop までの近傍イベント集合を使い, neighbor sampling [17] は 10 本とした. QIEA-TGX の観測操作は 2hop までのイベントを考慮し, 使用する量子ビット個体を 1 個, 回転行列と H ゲートのパラメータを $\theta_Q = 0.01\pi, \theta_H = 0.01$ とした. また, QIEA-TGX+において, ARGO 機構における回転角の最大値 θ_{max} と最小値 θ_{min} は, $\theta_{max} = 0.05\pi, \theta_{min} = 0.001\pi$, スケーリング係数は $K = 1.0$ とした. 局所探索における閾値

表 2: fidelity (Movielens, Enron)

	事実説明				反実仮想説明			
	Movielens		Enron		Movielens		Enron	
	TGAT	TGN	TGAT	TGN	TGAT	TGN	TGAT	TGN
random	0.083 \pm 0.12	0.003 \pm 0.15	0.189 \pm 0.11	0.075 \pm 0.16	0.239 \pm 0.20	0.171 \pm 0.15	0.131 \pm 0.13	0.262 \pm 0.20
GreeDy	0.099 \pm 0.12	0.023 \pm 0.16	0.145 \pm 0.15	0.082 \pm 0.17	0.366 \pm 0.22	0.228 \pm 0.17	0.312 \pm 0.20	0.355 \pm 0.25
PGExplainer	-0.080 \pm 0.18	-0.090 \pm 0.14	-0.098 \pm 0.22	-0.104 \pm 0.24	0.069 \pm 0.17	0.081 \pm 0.13	0.068 \pm 0.13	0.100 \pm 0.17
T-GNNEExplainer	0.052 \pm 0.17	-0.013 \pm 0.16	0.173 \pm 0.11	0.098 \pm 0.12	0.172 \pm 0.21	0.160 \pm 0.15	0.100 \pm 0.12	0.235 \pm 0.17
GA-TGX	0.132 \pm 0.10	0.059 \pm 0.12	0.188 \pm 0.11	0.071 \pm 0.17	0.226 \pm 0.12	0.125 \pm 0.08	0.106 \pm 0.11	0.254 \pm 0.19
QIEA-TGX	0.181 \pm 0.11	0.087 \pm 0.12	0.199 \pm 0.11	0.130 \pm 0.13	0.414 \pm 0.16	0.256 \pm 0.12	0.348 \pm 0.20	0.419 \pm 0.26
QIEA-TGX+	0.185 \pm 0.11	0.088 \pm 0.07	0.199 \pm 0.09	0.131 \pm 0.10	0.413 \pm 0.28	0.258 \pm 0.14	0.349 \pm 0.12	0.424 \pm 0.07

表 3: fidelity (Wikipedia, Reddit)

	事実説明				反実仮想説明			
	Wikipedia		Reddit		Wikipedia		Reddit	
	TGAT	TGN	TGAT	TGN	TGAT	TGN	TGAT	TGN
random	-0.015 \pm 0.19	-0.023 \pm 0.15	0.155 \pm 0.14	0.079 \pm 0.15	0.297 \pm 0.19	0.163 \pm 0.13	0.204 \pm 0.14	0.256 \pm 0.20
GreeDy	0.002 \pm 0.20	-0.005 \pm 0.16	0.139 \pm 0.15	0.090 \pm 0.15	0.435 \pm 0.21	0.236 \pm 0.15	0.473 \pm 0.23	0.366 \pm 0.25
PGExplainer	-0.210 \pm 0.17	-0.103 \pm 0.13	-0.174 \pm 0.24	-0.095 \pm 0.23	0.168 \pm 0.18	0.090 \pm 0.12	0.113 \pm 0.14	0.110 \pm 0.19
T-GNNEExplainer	-0.032 \pm 0.19	-0.043 \pm 0.16	0.152 \pm 0.13	0.095 \pm 0.11	0.242 \pm 0.20	0.157 \pm 0.13	0.151 \pm 0.13	0.246 \pm 0.18
GA-TGX	0.111 \pm 0.10	0.047 \pm 0.11	0.161 \pm 0.14	0.071 \pm 0.15	0.223 \pm 0.16	0.121 \pm 0.08	0.183 \pm 0.13	0.248 \pm 0.18
QIEA-TGX	0.132 \pm 0.11	0.074 \pm 0.11	0.179 \pm 0.14	0.123 \pm 0.12	0.456 \pm 0.23	0.253 \pm 0.16	0.490 \pm 0.22	0.421 \pm 0.25
QIEA-TGX+	0.134 \pm 0.13	0.075 \pm 0.08	0.180 \pm 0.06	0.122 \pm 0.11	0.465 \pm 0.13	0.256 \pm 0.12	0.497 \pm 0.18	0.430 \pm 0.05

τ_{low}, τ_{high} は, $\tau_{low} = 0.1, \tau_{high} = 0.9$ とした.

5.2 評価指標

評価指標には, fidelity [31] および実行時間を使用する. fidelity は, 説明として抽出された部分グラフが予測結果に与える影響度 (忠実度) を測定する指標であり, 以下のように定義される.

$$\mathcal{F}(f(\cdot), \mathcal{G}^c, \mathcal{G}^e, e_k) = \mathbb{1}(Y_k = 1)(f(\mathcal{G}^e)[e_k] - f(\mathcal{G}^c)[e_k]) + \mathbb{1}(Y_k = 0)(f(\mathcal{G}^c)[e_k] - f(\mathcal{G}^e)[e_k])$$

反実仮想説明においては, 現在のグラフ \mathcal{G}^c から説明部分グラフ \mathcal{G}^e を除外した場合の影響度として, 以下のように定義される.

$$\mathcal{F}(f(\cdot), \mathcal{G}^c, \mathcal{G}^e, e_k) = \mathbb{1}(Y_k = 1)(f(\mathcal{G}^c)[e_k] - f(\mathcal{G}^c \setminus \mathcal{G}^e)[e_k]) + \mathbb{1}(Y_k = 0)(f(\mathcal{G}^c \setminus \mathcal{G}^e)[e_k] - f(\mathcal{G}^c)[e_k])$$

ここで $\mathcal{F}(\cdot)$ は個別のイベントに対する fidelity であり, 実験の評価指標としては, 各イベント e_k の $\mathcal{F}(\cdot)$ の平均をとったものを用いる. fidelity は $[-1, 1]$ の範囲で値をとり, 値が大きい方が望ましい.

5.3 実験結果

5.3.1 fidelity の評価

表 2 に Movielens と Enron, 表 3 に Wikipedia と Reddit の実行結果を示す. 各 fidelity の値は, 存在するイベント, 存在しないイベントに対する説明の平均値である. ベースライン手法と比較して, 全 16 パターンの実験設定中 13 パターンにおいて最も高い fidelity を達成した. 四つのデータセット全

てに対して有効性が確認されており, 汎用性の高い手法と言える. QIEA-TGX と比較すると, 同等, もしくはわずかに上回る fidelity を記録した. 反実仮想説明における Wikipedia, Reddit など, 特定のパターンでは比較的大きなスコア上昇が確認できる.

5.3.2 fidelity と 実行時間の評価

図 3 に, ベースライン手法との fidelity と 実行時間の比較を示す. データセット Movielens と Enron で比較し, 各グラフで TGAT, TGN の結果をそれぞれプロットしている. QIEA-TGX+ は, 全てのグラフで QIEA-TGX と比べて点がわずかに右寄りに位置しており, QIEA-TGX よりも実行時間が長いことが分かる. これは局所探索機構による影響であり, QIEA による探索で絞り込めなかったイベントが多いと, 探索すべき組み合わせも増大してしまい, 実行時間に影響を及ぼすことを示している.

5.3.3 アブレーションスタディ

表 4 に Movielens, Enron, 表 5 に Wikipedia, Reddit における, ARGO と局所探索それぞれの機構のみを導入した場合の fidelity 変化を示す. ほとんどのパターンでは提案手法が最も高い fidelity を記録しているが, 反実仮想説明の Movielens(TGAT) では局所探索のみを導入した場合の方が fidelity が高いことが確認できる. これは ARGO のみを導入した場合が QIEA-TGX よりも fidelity が低いためであると考えられる. ARGO は多くの場合, 動的な回転角調整によりグラフに応じた最適な収束を促進するが, 促進が強すぎると探索の多様性が失われ, 探索空間の広い範囲を調べる前に質の低い局所解に収

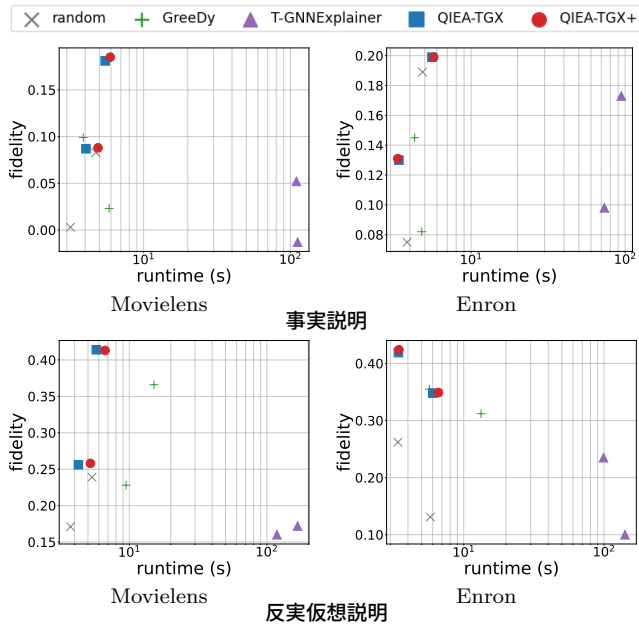


図 3: fidelity と実行時間

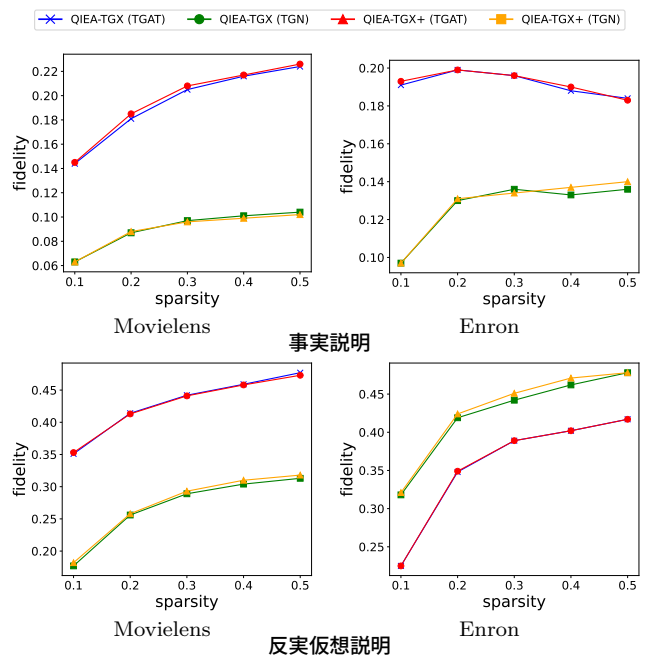


図 4: fidelity と sparsity

表 4: アブレーションスタディ (MovieLens, Enron)

	事実説明				反実仮想説明			
	MovieLens		Enron		MovieLens		Enron	
	TGAT	TGN	TGAT	TGN	TGAT	TGN	TGAT	TGN
QIEA-TGX	0.181	0.199	0.087	0.130	0.414	0.348	0.256	0.419
ARGO のみ	0.181	0.198	0.084	0.130	0.409	0.340	0.256	0.420
局所探索のみ	0.182	0.199	0.087	0.130	0.416	0.348	0.256	0.419
QIEA-TGX+	0.185	0.199	0.088	0.131	0.413	0.349	0.258	0.424

表 5: アブレーションスタディ (Wikipedia, Reddit)

	事実説明				反実仮想説明			
	Wikipedia		Reddit		Wikipedia		Reddit	
	TGAT	TGN	TGAT	TGN	TGAT	TGN	TGAT	TGN
QIEA-TGX	0.132	0.047	0.179	0.123	0.456	0.253	0.490	0.421
ARGO のみ	0.133	0.074	0.179	0.122	0.456	0.252	0.489	0.421
局所探索のみ	0.132	0.075	0.179	0.123	0.465	0.256	0.496	0.429
QIEA-TGX+	0.134	0.075	0.180	0.122	0.465	0.256	0.497	0.430

束してしまう可能性がある。したがって、局所探索等の機構で ARGO による早期収束を補正する必要がある。実際に、ARGO のみ、局所探索のみでは QIEA-TGX と比べて fidelity が上昇しない場合でも、両方の機構を導入することによって fidelity が上昇するパターンがいくつか存在する。

5.3.4 sparsity による fidelity 変化

図 4 に、sparsity を 0.1, 0.2, 0.3, 0.4, 0.5 と変化させたときの fidelity 推移を示す。ほとんどのパターンで QIEA-TGX+ は QIEA-TGX よりも高い fidelity を達成しており、sparsity の変化に対しても堅牢な手法だと言える。事実説明における Enron (TGAT) のグラフでは sparsity が上がると fidelity が下がってしまっているが、これは選択可能なイベント数増加により探索するイベントが増え、大域最適解に行きつかなくなったためだと考えられる。

6 おわりに

本研究では TGNN の説明手法である QIEA-TGX のさらなる説明精度向上を目的とし、QIEA-TGX に対し「ARGO による適応的更新」と「局所探索による事後最適化」の二つの改良機構を導入した手法 QIEA-TGX+ を提案した。実験の結果、提案手法は全実験設定の平均において、既存のベースライン手法を上回る説明精度を達成した。

謝 辞

本研究は ASPIRE JPMJAP2328 および JSPS 科研費 JP23K28096 の支援によって行われた。

文 献

- [1] Jiafeng Xiong, Ahmad Zareie, and Rizos Sakellariou. A survey of link prediction in temporal networks. *SN Computer Science*, 7(1):100, 2026.
- [2] Peng Wang, Baowen Xu, Yurong Wu, and Xiaoyu Zhou. Link prediction in Social Networks: The state-of-the-art. *Science China Information Sciences*, 58(1):1–38, 2015.
- [3] P. Borgnat, G. Dewaele, K. Fukuda, P. Abry, and K. Cho. Seven years and one day: Sketching the evolution of internet traffic. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, pages 711–719, 2009.
- [4] C. Vinchoff, N. Chung, T. Gordon, L. Lyford, and M. Aibin. Traffic prediction in optical networks using Graph Convolutional Generative Adversarial Networks. In *Proceedings of the International Conference on Transparent Optical Networks (ICTON)*, pages 1–4, 2020.
- [5] Chenxi Liu, Zhu Xiao, Dong Wang, Minhao Cheng, Hongyang Chen, and Jiawei Cai. Foreseeing private car transfer between urban regions with multiple graph-based Generative Adversarial Networks. *World Wide Web*,

- 25(2):2515–2534, 2022.
- [6] Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for Knowledge Graphs. *Proceedings of the IEEE*, 104(1):11–33, 2016.
 - [7] Meihong Wang, Linling Qiu, and Xiaoli Wang. A survey on Knowledge Graph Embeddings for link prediction. *Symmetry*, 13(3):485, 2021.
 - [8] Antonio Longa et al. Graph neural networks for temporal graphs: State of the art, open challenges, and opportunities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
 - [9] Wenwen Xia, Mincai Lai, Caihua Shan, Yao Zhang, Xinnan Dai, Xiang Li, and Dongsheng Li. Explaining Temporal Graph Models through an explorer-navigator framework. In *International Conference on Learning Representations (ICLR)*, 2023.
 - [10] Jialin Chen and Rex Ying. TempME: Towards the explainability of Temporal Graph Neural Networks via motif discovery. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pages 29005–29028, 2023.
 - [11] Zhan Qu, Daniel Gomm, and Michael Färber. CoDy: Counterfactual explainers for Dynamic Graphs. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2025.
 - [12] Gexiang Zhang. Quantum-Inspired Evolutionary Algorithms: A survey and empirical study. *Journal of Heuristics*, 17(3):303–351, 2011.
 - [13] Masahiro Mitani and Yuya Sasaki. Explaining Temporal Graph Neural Networks via Quantum-Inspired Evolutionary Algorithm. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2026.
 - [14] Lubna Siddiqui, Ashish Mani, and Jaspal Singh. Adaptive Quantum-Inspired Evolution for denoising PCG signals in unseen noise conditions. *IEEE Access*, 13:106188–106200, 2025.
 - [15] Dongsheng Luo, Wei Cheng, Dongkuan Xu, Wenchao Yu, Bo Zong, Haifeng Chen, and Xiang Zhang. Parameterized explainer for Graph Neural Network. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 19620–19631, 2020.
 - [16] Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. Inductive representation learning on Temporal Graphs. In *International Conference on Learning Representations (ICLR)*, 2020.
 - [17] Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael Bronstein. Temporal graph networks for deep learning on dynamic graphs. In *ICML Workshop on Graph Representation Learning*, 2020.
 - [18] Haonan Yuan, Qingyun Sun, Zhaonan Wang, Xingcheng Fu, Cheng Ji, Yongjian Wang, Bo Jin, and Jianxin Li. Dg-mamba: Robust and efficient dynamic graph structure learning with selective state space models. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 39, pages 22272–22280, 2025.
 - [19] Zeyang Zhang, Xin Wang, Ziwei Zhang, Haoyang Li, Yijian Qin, and Wenwu Zhu. LLM4DyG: Can large language models solve spatial-temporal problems on Dynamic Graphs? In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 4350–4361, 2024.
 - [20] Jaykumar Kakkad, Jaspal Jannu, Kartik Sharma, Charu C. Aggarwal, and Sourav Medya. A survey on explainability of graph neural networks. *IEEE Data Engineering Bulletin*, 46(2):33–44, 2023.
 - [21] 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 (NeurIPS)*, volume 32, pages 9637–9647, 2019.
 - [22] Ana Lucic, Maartje A Ter Hoeve, Gabriele Tolomei, Maarten De Rijke, and Fabrizio Silvestri. CF-GNNExplainer: Counterfactual explanations for Graph Neural Networks. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 4499–4511, 2022.
 - [23] David Bechtoldt and Sidney Bender. Graph diffusion counterfactual explanation. In *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 2026.
 - [24] Yinhan He, Zaiyi Zheng, Patrick Soga, Yaozhen Zhu, Yushun Dong, and Jundong Li. Explaining Graph Neural Networks with Large Language Models: A counterfactual perspective for molecular property prediction. In *Findings of the Association for Computational Linguistics: EMNLP*, pages 7079–7096, 2024.
 - [25] Kuk-Hyun Han and Jong-Hwan Kim. Quantum-Inspired Evolutionary Algorithm for a class of combinatorial optimization. *IEEE Transactions on Evolutionary Computation*, 6(6):580–593, 2002.
 - [26] Rohit Goyal, Krishan Kumar, Vivek Sharma, Rudramani Bhutia, Arpit Jain, and Munish Kumar. Quantum-inspired optimization algorithms for scalable machine learning in edge computing. In *Proceedings of the International Conference on Technological Advancements in Computational Sciences (ICTACS)*, pages 1888–1892, 2024.
 - [27] Mazyar Taghavi, Amir Paslari, and Mahdi Riazi. Quantum-Inspired Evolutionary Algorithm for hyperparameter tuning in automatic number plate recognition systems. In *Proceedings of the National Conference on Artificial Intelligence and Emerging Research*, 2025.
 - [28] Vadlamani Ravi and Pisipati Radha Krishna. Quantum-Inspired Evolutionary Algorithms for feature subset selection: A comprehensive survey. *Quantum Information Processing*, 24(7), 2025.
 - [29] R. Hinterding. Representation, constraint satisfaction and the knapsack problem. In *Proceedings of the Congress on Evolutionary Computation (CEC)*, volume 2, pages 1286–1292, 1999.
 - [30] Michael A Nielsen and Isaac L Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.
 - [31] Hao Yuan, Haiyang Yu, Shurui Gui, and Shuiwang Ji. Explainability in Graph Neural Networks: A taxonomic survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5782–5799, 2023.