

# 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<sup>†</sup> 道瀬 悠磨<sup>†</sup> 原 隆浩<sup>†</sup>

<sup>†</sup> 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

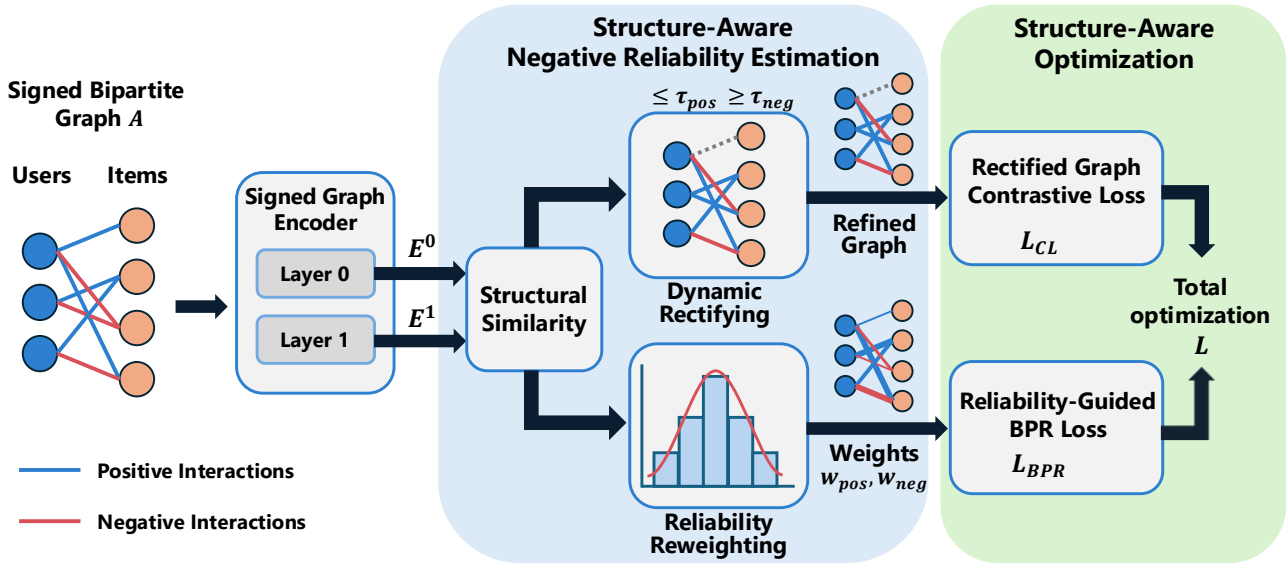


図 1: SANE の全体アーキテクチャ。

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、 $\tau$  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、 $\Theta$  はすべての学習可能パラメータを表し、 $\lambda_{CL}$  および  $\lambda_{reg}$  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
	<b>SANE (Ours)</b>	<b>0.1490*</b>	<b>0.0879*</b>	<b>0.3222*</b>	<b>0.1880*</b>	<b>0.1563*</b>	<b>0.0755*</b>
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  $\rho_{pos}$ , 偽陰性修正率  $\rho_{neg}$ , および対照学習の重み  $\lambda_{CL}$  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

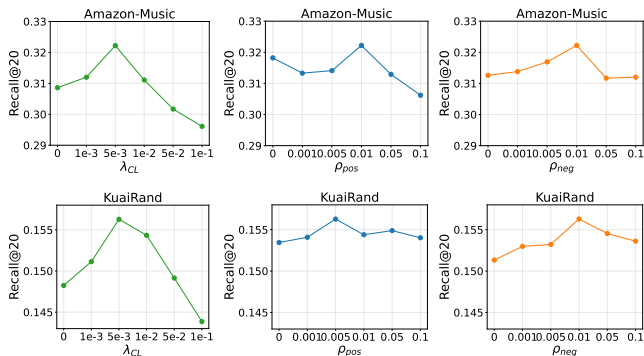
表 2 に全手法の性能比較結果を示す. SANE はすべてのデータセットにおいて全ベースラインを有意に上回り, Amazon-Music では Recall@20 で最大 6.34% の改善を達成した. これらの結果から, 以下の知見が得られた.

**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)	<b>0.3222</b>	<b>0.1880</b>	<b>0.1562</b>	<b>0.0755</b>

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

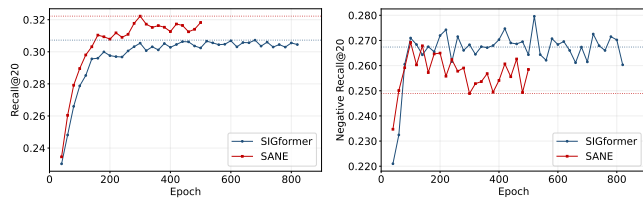
間を適応的に保持している。

**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  $\lambda_{CL}$ :**  $\lambda_{CL}$  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  $\lambda_{CL}$  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  $\rho_{pos}$  および  $\rho_{neg}$ :** 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  $\rho_{pos}$  が Amazon-Music で 0.01, KuaiRand で 0.005 であったのに対し,  $\rho_{neg}$  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,  $\hat{\mathcal{I}}_u$  はユーザー  $u$  に対して生成された Top-20 推薦リスト,  $\mathcal{E}_u^-$  はユーザー  $u$  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

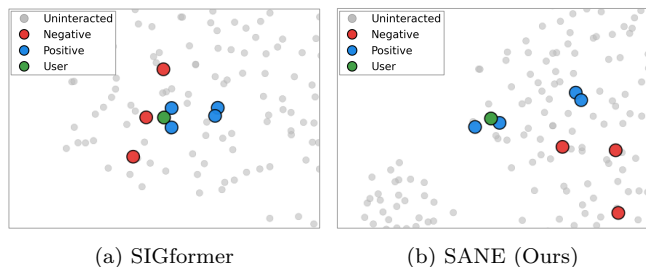


図 4: 学習済み埋め込み空間の t-SNE 可視化

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.