

レビューテキストと知識グラフに基づく推薦における説明性能の向上

Improving Explainability of Recommender Systems based on Review Texts and the Knowledge Graph

尾崎 健吾[†] 伊東 聖矢[‡] 大原 剛三[‡]
Kengo Ozaki Seiya Ito Kouzou Ohara

1 はじめに

現在、電子商取引やソーシャルネットワークサービス等のさまざまなサービスにおいて推薦システムが利用されている。推薦システムとは、ユーザの行動履歴やアイテムの情報を利用することによって、ユーザが好むアイテムを予測し、推薦するシステムである。これにより、ユーザは膨大なアイテムの中から、好みのアイテムを簡単に見つけることが可能になる。近年では、ユーザに推薦アイテムだけではなく、なぜその推薦アイテムがおすすめなのかという推薦理由を提示する説明可能な推薦の研究が行われている。説明可能な推薦により、推薦システムの透明性、説得力、有効性、信頼性、ユーザ満足度を向上させることができる [5]。

そのような説明可能な推薦において、知識グラフを用いるアプローチが数多く提案されている [4, 1, 3]。知識グラフとは、現実世界に存在する事象 (entity) とそれらの関係 (relation) をグラフ構造で表現したものである。推薦分野において、知識グラフをアイテムの補助情報として活用することは、推薦したアイテムに対する直感的な推薦理由を生成するのに役立つと考えられている [2]。知識グラフを用いた説明可能な推薦のアプローチのひとつとして、Wang ら [3] は Knowledge Graph Attention Network (KGAT) と呼ばれる手法を提案している。KGAT では、ユーザとユーザの購買アイテムで定義されるユーザ-アイテム間二部グラフと知識グラフを組み合わせた協調知識グラフと呼ばれるグラフ構造を学習することで、近傍ノードに基づいた各ノードに対するベクトル表現、および、各ノード間の関連性の強さをエッジの重みで表現した重み付き有向グラフを獲得する。そして、学習の結果得られるエッジの重みを解析することにより、たとえば、「過去に購入したアイテムと同じジャンルであるから」「同じアイテムを購入した他のユーザがこのアイテムを購入しているから」などの推薦理由を提示することを可能にしている。

KGAT を筆頭とした知識グラフを利用する推薦手法では、高い推薦精度と推薦理由を示すことを可能としている一方で、ユーザの購買アイテムが少ない場合、もしくはアイテムが知識グラフ上のエンティティと繋がりが少ない場合などでは、推薦に用いるユーザ-アイテムのベクトル表現を適切に学習できない場合がある。ベクトル表現を適切に獲得できれば、推薦精度の向上が見込める。また、推薦理由として提示する知識グラフの情報は各ユーザの嗜好に応じたものではないため、ユーザが購買アイテムに対して書いたレビューテキストに含まれるユーザの明示的嗜好もしくはアイテムの特徴を表す情報を選択的に利用することができれば、パーソナライズされた推薦理由の網羅性向上が期待できる。

そこで本研究では、レビューテキストを用いて知識グラフを拡張することで、推薦精度と説明性能の向上を目指す。具体的には、レビューテキストからユーザの明示的嗜好もしくはアイテムの特徴を示す特徴語を抽出し、協調知識グラフを拡張する。提案手法を適用することで、推薦理由として知識グラフ上のエンティティや他のユーザを提示するだけでなく、ユーザの明示的嗜好もしくはアイテムの特徴を示す特徴語を提示することが可能となり、より具体的でパーソナライズされた推薦理由の生成が可能となる。評価実験では、提案手法により拡張した知識グラフを KGAT の入力とし、推薦精度および推薦における説明性能を定量的に評価する。

2 提案手法

2.1 提案手法の概要

提案手法では、従来の KGAT をユーザの明示的嗜好情報とアイテムの特徴を考慮したモデルへと拡張する。具体的には、レビューテキストからユーザの明示的嗜好とアイテムの特徴を示す特徴語を抽出し、ユーザと特徴語、アイテムと特徴語のペアで知識グラフ上の triplet を定義する。このようにユーザ-アイテム間で共有可能な特徴語を用いて協調知識グラフを拡張することで、ユーザ-アイテム間のつながりを強固にし、推薦精度ならびに説明性能の向上を図る。

2.2 協調知識グラフの拡張

2.2.1 特徴語集合の抽出

レビューテキストから特徴語を抽出する際には、感情分析ツールを利用し、(特徴語, 感情極性)=(F , S) の組を抽出する。 (F, S) の具体例として、(“character”, +1) や (“story”, -1) といったものが挙げられる。ここで、+1 はポジティブな感情極性、-1 はネガティブな感情極性を表す。ここで得られる特徴語 F の表記揺れ等を解消し、出現頻度の高い単語の中からユーザの明示的嗜好もしくはアイテムの特徴として不適切な単語を除去したものを特徴語集合 *feature_words* として定義する。

2.2.2 triplet の生成

前述のように定義した特徴語集合 *feature_words* を用いて、ユーザとそのユーザが過去に言及した特徴語の間でユーザ-特徴語 triplet (UF: User-Feature triplet)、および、アイテムとそのアイテムについてのレビューテキスト内で述べられている特徴語の間でアイテム-特徴語 triplet (IF: Item-Feature triplet) を定義する。提案手法では、ユーザと特徴語間、アイテムと特徴語間の関係をそれぞれ“好む”、“好まれる”とする。ユーザ-特徴語 triplet の生成に関しては、ユーザ u とそのユーザ u が過去に言及したレビューに含まれる特徴語 F_u が「感情極性が +1」かつ「全体のレビューから定義した *feature_words* に含まれる」という条件を満たす場合に限り、ユーザ-特徴語 triplet(u , “好む”, F_u) を定義する。これにより、ユーザと、そのユーザが好む特徴語のみで

[†] 青山学院大学大学院理工学研究科

[‡] 青山学院大学理工学部

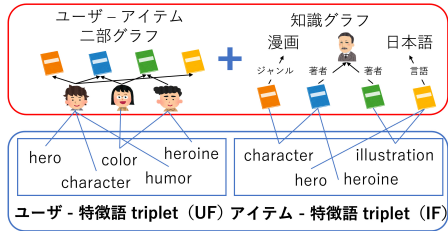


図 1 拡張知識グラフのイメージ

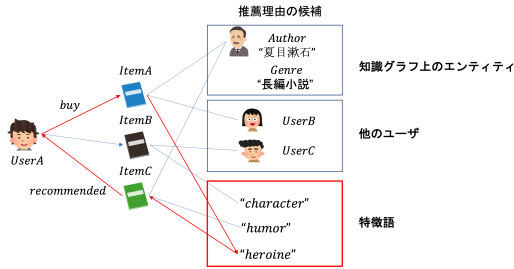


図 2 提案手法を適用した KGAT により推薦されたアイテムの推薦理由の解釈

triplet を定義することができ、ユーザの明示的嗜好情報を捉えることができる。アイテム-特徴語 triplet の生成に関しては、アイテム i とそのアイテム i の特徴語 F_i が「アイテム i のレビュー集合から抽出できる F_i の感情極性の和が +1 以上」かつ「全体のレビューから定義した $feature_words$ に含まれる」の条件を満たす場合に限りアイテム-特徴語 triplet ($i, \text{“好まれる”}, F_i$) を定義する。これにより、アイテムと多くのユーザがポジティブに評価しているそのアイテムの特徴語で triplet を定義することができる。

そして、ユーザとユーザの購買アイテムで定義されるユーザ-アイテム間二部グラフをユーザ-特徴語 triplet で拡張し、知識グラフをアイテム-特徴語 triplet で拡張することで最終的な拡張知識グラフを生成する。生成する拡張知識グラフのイメージを図 1 に示す。

図 2 は、提案手法を適用した KGAT により推薦されたアイテムの推薦理由の解釈の例である。拡張前の知識グラフを用いた推薦では、Author や Genre 等の拡張前の知識グラフに含まれる情報と他のユーザしか推薦理由として挙げることができないが、提案手法によりユーザの明示的な嗜好情報やアイテムの特徴を示す“character”, “humor”, “heroine”といった特徴語を推薦理由として提示することも可能になる。

3 評価実験

3.1 実験設定

本実験では、提案手法と既存手法における推薦精度と説明精度を評価する。提案手法では各ユーザに提示する推薦理由の候補が、特徴語、他のユーザ、拡張知識グラフ中のエンティティの 3 種類であるため、説明精度の定量的な評価にはそれぞれ異なる評価指標を使用する必要がある。ただし、拡張知識グラフ中のエンティティが推薦理由となる場合は正解データを用意できないため、定量的な評価は困難である。そのため、特徴語が推薦理由となる場合、他のユーザが推薦理由となる場合の 2 つの場合についてのみ、それぞれ推薦精度と説明精度を求めた。

3.1.1 評価指標

推薦精度の評価には、 $Recall@N$ を採用した。説明精度の評価に関しては、特徴語が推薦理由となる場合と他のユーザが推薦理由となる場合では異なる評価指標を用いた。具体的には、前者に関しては、推薦理由となる特徴語が実際に被推薦ユーザが好む特徴語リストに含まれている確率 $explainability$ を定義し、全ユーザの平均を最終的な説明精度とした。 $explainability$ は、式 (1) のように定義する。

$$explainability = \frac{1}{|U'|} \sum_{u' \in U'} \frac{count}{|I_{u', r_i}|} \times 100 \quad (1)$$

$$count = \sum_{i \in I_{u'}} reason_i \quad (2)$$

$$reason_i = \begin{cases} 1 & \text{if } r_i \text{ in } F_{u'} \\ 0 & \text{else} \end{cases} \quad (3)$$

ここで、 U' はレビューから特徴語が抽出できたユーザの集合、 r_i はアイテム i に対する推薦理由、 I_{u', r_i} はユーザ $u' \in U'$ が好むアイテムのうち r_i が特徴語に含まれるものの集合、 $F_{u'}$ はユーザ u' の過去のレビューから抽出した特徴語リストとする。

一方、他のユーザが推薦理由となる場合に関しては、被推薦ユーザと推薦理由となるユーザの嗜好の類似度をジャカード係数 ($jaccard_sim$)、ダイス係数 ($dice_sim$) を用いて評価した。それぞれ、式 (4) と (5) で定義される。

$$jaccard_sim = \frac{|F_a \cap F_b|}{|F_a \cup F_b|} \quad (4)$$

$$dice_sim = \frac{2 \times |F_a \cap F_b|}{|F_a| + |F_b|} \quad (5)$$

ここで F_a , F_b はそれぞれ被推薦ユーザ、推薦理由ユーザの好む特徴語集合である。

3.1.2 比較手法

本実験で比較する手法を以下に示す。

KGAT [3]

知識グラフを利用するベースライン手法。

KGAT+IF&UF

KGAT に IF と UF を用いて拡張した拡張知識グラフを入力する手法 (提案手法)

KGAT+IF

KGAT に IF を用いて拡張した拡張知識グラフを入力する手法 (提案手法の部分手法)

EFM [6]

行列因子分解によってアイテムの特徴に対するユーザの嗜好度およびアイテムがもつ特徴の重要度を予測し、ユーザがアイテムに対して好む特徴を説明する手法。

実験では、特徴語が推薦理由となる場合の KGAT+IF、KGAT+IF&UF に対するベースラインとして EFM を用い、他のユーザが推薦理由となる場合の KGAT+IF のベースラインとして KGAT を用いた。

3.1.3 データセット

本実験では、Amazon においてレビューを 5 件以上投稿しているユーザが対象となっている 5-Core データ

表 1 特徴語が推薦理由となる場合の実験のデータセット概要

	Amazon Books	EFM	KGAT+IF	KGAT+IF&UF
ユーザ-アイテム間 二部グラフ	Users		6,504	6,504
	Items		5,441	5,441
	Feature words		488	488
	Interactions		150,274	184,240
知識グラフ	Entities	-	89,060	
	Relations	-	40	
	triplets	-	2,077,696	

表 2 他のユーザが推薦理由となる場合の実験のデータセット概要

	Amazon Books	KGAT	KGAT+IF
ユーザ-アイテム間 二部グラフ	Users	70,679	
	Items	24,915	
	Interactions	847,333	
知識グラフ	Entities	88,572	89,060
	Relations	39	40
	triplets	2,557,746	2,638,917
	Feature words	-	488

※ Interactions はユーザ-アイテム間のエッジの数を示す

セット¹ (以下, Amazon Dataset) の Books カテゴリのデータを使用した. Amazon Dataset では, ユーザがアイテムに対して付ける評価値に加えて, 意見や感想などを含むレビューテキストが付随する. 特徴語の抽出には, レビューテキストに対して感情分析ツールである Sentires² [7] を適用し, (特徴語, 意見, 感情極性)=(F , O , S) の 3 字組が抽出可能な場合, 抽出した特徴語をユーザが好む特徴語およびアイテムの特徴語とみなし, ユーザ-特徴語 triplet (UF), アイテム-特徴語 triplet (IF) を生成した. 特徴語集合 *feature_words* は, 抽出した特徴語の出現頻度上位 25%から選定し, 最終的に 488 単語とした. 入力として用いる知識グラフは, Zhao らの研究 [8] に従って作成した.

異なる 2 つの実験に対するデータセットの概要をそれぞれ表 1 と表 2 に示す. 表 1 は, 特徴語が推薦理由となる場合の実験に対するデータセットの概要である. EFM は Amazon Dataset のうち特徴語が抽出可能なレビューをもつアイテム (5,441 個) のみしか用いることが出来ないため, それに合わせてデータセットを作成した. ユーザ-アイテム間二部グラフに関しては, EFM と KGAT+IF で同じであるが, KGAT+IF&UF ではユーザ-特徴語間で triplet を定義しているため, その分 Interactions が多くなっている. また, 知識グラフに関しては, KGAT+IF と KGAT+IF&UF で共通である. 一方, 表 2 は, 他のユーザが推薦理由となる場合の実験に対するデータセットの概要である. こちらに関しては, Amazon Dataset のうち, 特徴語が抽出できないレビューしかもたないアイテムも含めて作成した. ユーザ-アイテム間二部グラフに関しては, KGAT と KGAT+IF で共通である. これに対して知識グラフに関しては, KGAT+IF ではアイテム-特徴語間で triplet を定義しているため, その分 Entities, Relations, triplets の数が多くなっている.

¹ <http://jmcauley.ucsd.edu/data/amazon/>

² <https://github.com/evison/Sentires>

表 3 特徴語が推薦理由となる場合の推薦精度と説明精度

推薦手法	Recall@10	explainability
EFM	0.0549	37.55
KGAT+IF	0.0509	52.28
KGAT+IF&UF	0.0500	83.84

表 4 他のユーザが推薦理由となる場合の推薦精度と説明精度

推薦手法	Recall@20	Jaccard Sim	Dice Sim
KGAT	0.150	0.167	0.237
KGAT+IF	0.151	0.178	0.246

3.1.4 学習設定

KGAT, KGAT+IF, KGAT+IF&UF の学習設定は, Wang ら [3] の研究に従った. まず, データを訓練データとテストデータに 4:1 の比率で分割し, ハイパーパラメータの調整には, 訓練データの 10%を検証用データとして用いた. モデルの重みの初期化には Xavier 初期化を用い, Embedding size は 64 で固定した. モデルの最適化に関しては Adam Optimizer を使用し, その際のバッチサイズは 1,024 で固定し, 学習率は 0.001, ノードの Dropout 率は 0.1 とした. 学習の際, Early stopping を設定し, 50 エポックの間, 検証用データで Recall@N の値が上がりななければ学習を終了し, その時点のエポックまでで最も精度の高いものを最終的な精度とした.

3.2 評価結果

3.2.1 特徴語が推薦理由となる場合

表 3 に, 特徴語が推薦理由となる場合の推薦精度と説明精度を示す. 推薦精度に関しては EFM が KGAT+IF, +IF&UF よりも約 0.0040 程高く, 最も高い精度となった. 一方, 説明精度に関しては, KGAT+IF, +IF&UF が EFM を上回り, 特に KGAT+IF&UF は EFM を約 46.3 ポイント上回った. KGAT+IF&UF によって高い説明精度を獲得できた原因としては, KGAT+IF との比較から, ユーザ-特徴語間で triplet を定義したことの影響が大きいと考えられる. アイテム-特徴語間のみでなく,

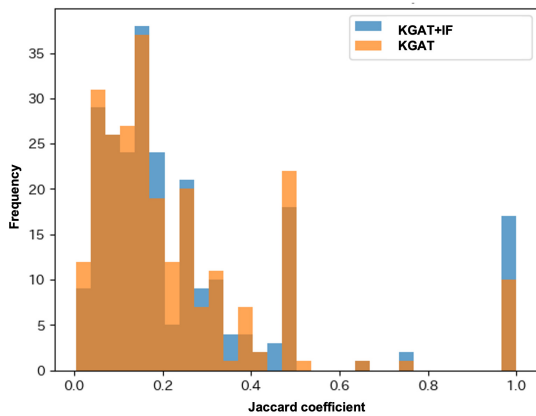


図 3 被推薦ユーザと推薦理由ユーザの特徴語集合に対するジャッカード係数の分布

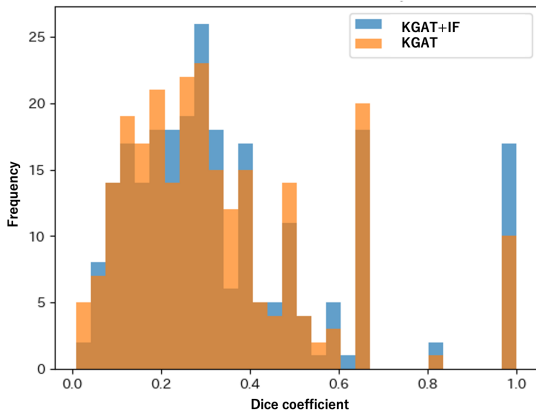


図 4 被推薦ユーザと推薦理由ユーザの特徴語集合に対するダイス係数の分布

ユーザ-特徴語間で triplet を定義することで、ユーザと特徴語の埋め込み表現が近くなるように学習され、両者の間の注意重みが大きくなり、ユーザの好む特徴語を推薦理由として適切に提示することができたと考えられる。

3.2.2 他のユーザが推薦理由となる場合

他のユーザが推薦理由となる場合の実験結果を表 4 に示す。Recall@20 による推薦精度の評価に関しては、拡張知識グラフを用いた KGAT+IF の方が、拡張前の知識グラフを用いる KGAT よりも高い精度となっているものの、拡張知識グラフを用いることによる推薦精度の大幅な向上は確認はできず、期待していた結果を得ることはできなかった。説明精度に関しては、入力として拡張知識グラフを用いた KGAT+IF の方が、拡張前の知識グラフを用いる KGAT よりもジャッカード係数で約 6.59%、ダイス係数で約 3.80%程の類似度の向上が確認できる。図 3 と図 4 は、被推薦ユーザと推薦理由ユーザの特徴語に基づく類似度の分布を示したものである。これらの図からは、ジャッカード係数、ダイス係数とも

に、類似度が 0.2 以下となる被推薦ユーザと推薦理由ユーザのペア数に関しては、拡張前の知識グラフを用いる KGAT のほうが拡張知識グラフを用いる KGAT+IF より多くなっており、その一方で、類似度が 0.2 以上、特に 1.0 付近となるペア数に関しては、拡張知識グラフを適用した KGAT+IF の方が拡張前の知識グラフを利用する KGAT よりもその数は多くなっている。言い換えると、拡張知識グラフを用いる KGAT+IF の方が、被推薦ユーザと多くの特徴語を共有する、より類似度の高い他のユーザを推薦理由として選択できているといえる。

4 おわりに

本研究では、推薦システムの推薦精度と説明性能の向上のために、従来手法である KGAT をユーザの明示的な嗜好情報とアイテムの特徴を考慮したモデルへと拡張した。評価実験の結果、提案手法が説明性能の向上に有効であることを確認できた一方で、推薦精度には大きく寄与しないことを確認した。今後の課題として、ユーザの言及回数に応じた特徴語の重み付けなどによる、より詳細な推薦理由の生成が挙げられる。

参考文献

- [1] Qingyao Ai, Vahid Azizi, Xu Chen, and Yongfeng Zhang. Learning heterogeneous knowledge base embeddings for explainable recommendation. *Algorithms*, 11(9):137, 2018.
- [2] Janneth Chicaiza and Priscila Valdiviezo-Diaz. A comprehensive survey of knowledge graph-based recommender systems: Technologies, development, and contributions. *Information*, 12(6):232, 2021.
- [3] 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 (KDD)*, pages 950–958, 2019.
- [4] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, page 353–362, 2016.
- [5] Yongfeng Zhang and Xu Chen. Explainable recommendation: A survey and new perspectives. *Foundations and Trends in Information Retrieval*, 14(1):1–101, 2020.
- [6] Yongfeng Zhang, Guokun Lai, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR)*, pages 83–92, 2014.
- [7] Yongfeng Zhang, Haochen Zhang, Min Zhang, Yiqun Liu, and Shaoping Ma. Do users rate or review? boost phrase-level sentiment labeling with review-level sentiment classification. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR)*, pages 1027–1030, 2014.
- [8] Wayne Xin Zhao, Gaole He, Kunlin Yang, Hongjian Dou, Jin Huang, Siqi Ouyang, and Ji-Rong Wen. Kb4rec: A data set for linking knowledge bases with recommender systems. *Data Intelligence*, 1(2):121–136, 2019.