

Knowledge Graph Attention Network に基づく顧客分析手法に関する研究

1X16C011-1 伊藤史世
指導教員 後藤正幸

1. 研究背景・目的

従来のマーケティング活動では、消費者を属性情報を用いてセグメンテーションし、セグメントごとに施策を検討するというアプローチが取られてきた。消費者行動には、性別や年代などの属性情報による類似性があり、このアプローチには一定の効果がある。すなわち、同一の属性を持つ消費者はある程度類似した嗜好（以下、ユーザ属性に依る嗜好）を持つといえる。これに対し、現在の EC サイトにおけるマーケティング施策を考えた場合、ユーザの購買行動を取得することが可能となったため、個々のユーザの嗜好にカスタマイズされた様々な施策が実施されている。そのような「ユーザ固有の嗜好」にカスタマイズし、成功を収めているマーケティング技術の1つに推薦システムがある。ここで、推薦システムとは、ユーザの購買履歴データなどを活用し、各ユーザの嗜好に合致したアイテムを推薦するシステムを指す。これまでの推薦システムは、個々のユーザの購買行動を詳細に分析し、「ユーザ固有の嗜好」を正確に推定しようとする方向で発展してきた。そのような方向性で発展した推薦モデルの1つとして、近年、Deep Learning をベースとした Knowledge Graph Attention Network（以下、KGAT）[1] が提案されている。このモデルでは、アイテムの補助情報を取り入れて、ユーザと購買アイテム、購買アイテムとその補助情報の関係性をモデル化し、それらの関係性の強さを定量化することができる。この定量化された関係性の強さを用いて、各ユーザに対するアイテムの推薦に解釈を与えることが可能となっている。そのため、KGAT は推薦のためのモデルのみならず、ユーザの購買行動分析に対して、何らかの解釈を与える分析モデルとして活用できる可能性がある。しかしながら、従来の KGAT は推薦を目的としたモデルであるため、アイテムの補助情報を推薦のために活用しているのみであり、ユーザの属性情報については考慮されていなかった。

そこで、本研究では、KGAT をユーザとその属性情報の関係性を新たに考慮したモデルへと拡張する。これにより、ユーザの嗜好をユーザの属性情報と購買アイテムを用いて解釈することが可能となる。提案手法を用いることで、「ユーザ属性に依る嗜好」と「ユーザ固有の嗜好」を統合的に分析することが可能になり、より柔軟なマーケティングの施策立案に役立つことが期待される。最後に、実際の EC サイトの評価履歴データに提案手法を適用し、分析結果を述べると共にその有用性を示す。

2. 従来モデル

KGAT は、ユーザと購買アイテム、購買アイテムの補助情報からなるグラフを入力として、ユーザの各アイテムに対する購買確率を出力する推薦モデルである。このモデルでは、グラフ上の高次の関係性を捉えながら、推薦において重要である関係を学習し、ユーザやアイテム、アイテムの補助情報

をベクトルで表現する。

3. 提案手法

3.1. 概要

Wang らの研究 [1] では、同じアイテムを購入したことのあるユーザ同士が購買アイテムを介してグラフ上で接続されており、その嗜好の類似性は購買アイテムからモデル化されていた。これに対し、提案手法ではユーザの属性情報をグラフに組み込むことを考える。すると、ユーザは購買アイテムのみでなく、属性情報を介して同一の属性をもつ他のユーザと接続されるようになり、ユーザの嗜好を属性情報と購買アイテムからモデル化することができる。ここでは、ユーザとその属性情報、アイテムとその属性情報である店舗・カテゴリ情報により構成されるグラフに対して KGAT を適用する。

3.2. グラフの構築

グラフは、ノード（頂点）集合、エッジ（辺）集合から構成される。提案モデルの入力とするグラフは、各ノードをユーザとアイテムなどの集合としてエンティティ、エッジを「購買する」や「〜で販売されている」などのリレーションと呼ぶ。ここで、グラフ上のエンティティを $q \in Q$ 、リレーションを $r \in R$ とする。提案モデルでは、このエンティティとしてユーザ集合 U 、アイテム集合 I 、アイテムの属性として店舗集合 A_1 、カテゴリ集合 A_2 に加えてユーザの属性集合 A_3 も考慮し、エンティティ集合を $Q = U \cup I \cup A_1 \cup A_2 \cup A_3$ とする。また、グラフ G は、「先頭エンティティ」、「リレーション」、「末尾エンティティ」の3つの組であるトリプル (h, r, t) の集合 $\mathcal{G} = \{(h, r, t) | h \in Q, r \in R, t \in Q\}$ として定義される。

3.3. モデル式

提案モデルではエンティティとリレーションをベクトルで表現する。 $p \in Q \cup R$ のベクトル表現を e_p とし、ベクトル表現の次元数を d とする。ここでは、TransR[2] に基づき、知識グラフ上のエンティティのベクトル表現 e_h, e_t を、リレーションごとに異なる意味空間上に写像し、リレーション $r \in R$ による写像行列を W_r とする。グラフのスコア関数 g は、 $g(h, r, t) = \|W_r e_h + e_r - W_r e_t\|_2$ として定義され、グラフ構造に対する損失は、式 (1) で表される。

$$Loss_{KG} = \sum_{(h,r,t,t') \in \mathcal{T}} -\ln \sigma(g(h, r, t') - g(h, r, t)) \quad (1)$$

ただし、 \mathcal{T}_r をリレーション r について定義されている末尾エンティティの集合として、 $\mathcal{T} = \{(h, r, t, t') | t, t' \in \mathcal{T}_r, (h, r, t) \in \mathcal{G}, (h, r, t') \notin \mathcal{G}\}$ である。さらに、グラフに存在するトリプル (h, r, t) のエッジの重み $\pi(h, r, t)$ を以下の式で求める。

$$\hat{\pi}(h, r, t) = (W_r e_t)^\top \tanh(W_r e_h + e_r) \quad (2)$$

$$\pi(h, r, t) = \frac{\exp(\hat{\pi}(h, r, t))}{\sum_{(h,r',t') \in \mathcal{N}_h} \exp(\hat{\pi}(h, r', t'))} \quad (3)$$

ただし、 \mathcal{N}_h は先頭エンティティ h に対してグラフ上で定義されているトリプルの集合である。

続いて、 $\pi(h, r, t)$ により、 e_h に対する周辺のエンティティのベクトル表現の重み付け和を計算し、その表現を自身に取り入れることを L 回繰り返すことにより、グラフ上の L 次近傍の関係性を考慮したベクトル表現を獲得する [1]。このとき、 e_h は $\pi(h, r, t)$ の大きい e_t に依存した表現となる。そのため、ユーザ $u \in \mathcal{U}$ の属性を $a_u \in \mathcal{A}_3$ としたとき、 $\pi(u, r, a_u)$ によって、ユーザと属性の関連度を定量化することが可能となる。

3.4. 提案モデルを用いた分析方法

ユーザ u が同一の属性情報を持つユーザ間で頻繁にみられる購買行動をしていた場合、 $\pi(u, r, a_u)$ は大きな値となり、属性に依る嗜好が強まると解釈することができる。一方、ユーザの購買行動が属性に依らず、固有の嗜好に基づいている場合、この値は小さくなる。また、ユーザ u が購買したアイテム i に関する重み $\pi(u, r, i)$ により、ユーザ固有の嗜好がどの購買アイテムにより説明可能であるかを議論することができる。

4. 実データを用いた分析

提案手法の有用性を示すため、2012年の楽天市場の評価履歴データ [3] に対して提案手法を適用し、分析を行う。

4.1. 分析条件

データ中の全ユーザ数は 27,641 人、全アイテム数は 123,751 個であった。ユーザ属性を性別 (男女) と年代 (10代-70代) の組み合わせで表現し、グラフのベクトル表現の次元数を $d = 64$ 、 $L = 2$ としてモデルの学習を行った。

4.2. 分析結果と考察

まず、各属性の嗜好の傾向に関する分析を行うために、各属性の $\pi(u, r, a_u)$ の平均値について比較を行う。性別で比較すると、その値は男性の方が全体的に低い傾向がある。このことから、男性はユーザ固有の嗜好が強くなり、女性は属性ごとに似た購買を行う傾向があるといえる。また、同一の性別における年代間の比較をすると、男性においては、30代、40代で大きくなり、年代が上昇するにつれて小さくなっている。一方で、女性は、年代が上昇するにつれて大きくなる傾向がある。このことから、男性は50代以降に、ユーザ固有の嗜好が強くなり、趣味などの個人の嗜好に基づく購買をするようになることが推測される。女性は、年代が上昇するにつれて属性との関連度が大きくなることから、その嗜好は均一化していく傾向にあると言える。

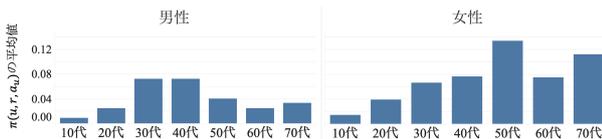


図 1: 属性ごとの $\pi(u, r, a_u)$ の平均値

さらに、 π の有用性について検討するため、 $\pi(u, r, a_u)$ と $\pi(u, r, i)$ について分析する。そこで、評価件数が 5 件のユーザのうち、 $\pi(u, r, a_u)$ が最も高かったユーザ u_1 、および最も低かったユーザ u_2 を抽出し、その購買傾向について考察する。該当するユーザの π の値を表 1 に示す。

表 1: 該当ユーザの購買履歴

u_1		u_2	
エンティティ名	π	エンティティ名	π
a_u :40 代男性	9.43×10^{-1}	ペット用パーカー	9.99×10^{-1}
生パスタセット	2.67×10^{-2}	ペット用ポロシャツ	2.35×10^{-8}
座いす	1.09×10^{-2}	ブランケット	7.40×10^{-9}
脚時計	8.01×10^{-3}	犬用サプリメント	2.42×10^{-9}
懐中電灯	6.82×10^{-3}	犬用デンタルケア	2.06×10^{-9}
DVD プレーヤー	5.03×10^{-3}	a_u :40 代女性	1.64×10^{-13}

表 1 より、ユーザ u_1 は日用品や食品などの複数のジャンルの商品を購入しており、偏った購買傾向はみられず、40代男性の標準的な購買行動をしているといえる。一方、ユーザ u_2 はペットに関する商品を多く購入しており、一貫したユーザ固有の嗜好に基づいた購買行動が確認できる。このように、 $\pi(u, r, a_u)$ の値が小さいユーザは、ユーザ固有の嗜好を確認することができる。

5. 考察

4 節の実データ分析の結果からも示されているように、提案モデルはユーザの嗜好を属性情報と購買アイテムの双方から説明可能なモデルとなっている。さらに、ユーザの属性情報への影響度 $\pi(u, r, a_u)$ が大きいユーザに対してはセグメントマーケティングを実施し、影響度が小さいユーザに対しては、ユーザにパーソナライズした施策を実施するなど、提案モデルはより柔軟なマーケティング戦略の考案にも有用であると考えられる。また、今回はユーザの嗜好に関する分析を行なったが、アイテムについても同様の分析は有効であると考えられる。アイテム i に関して、店舗 $s \in \mathcal{A}_1$ との関連度 $\pi(i, r, s)$ を算出することができる。この値を用いて、 $\pi(i, r, s)$ が高いアイテムは店舗に依存して購買されるというような、アイテムの性質に関する分析が可能となる。

以上により、ユーザやアイテムのどのような属性が購買行動に関係するのかを、本研究で定量的に明らかにすることが可能になったといえる。属性情報の与え方は柔軟に変化させることができるため、様々な分野に応用可能な手法である。

6. まとめと今後の課題

本研究では、ユーザの嗜好を属性と購買アイテムの双方の観点から解釈可能なモデルを提案し、実際の EC サイトの評価データを用いて提案手法の有用性を示した。今後の課題として、購買件数に基づく π の正規化アルゴリズムの考案や、購買順序の考慮といったモデルの改良が挙げられる。

謝辞：本研究では、国立情報学研究所の IDR データセット提供サービスにより楽天株式会社から提供を受けた「楽天データセット」を使用した。貴重なデータの提供に深く感謝致します。

参考文献

- [1] X. Wang, X. He, Y. Cao, M. Liu and T.S. Chua, "KGAT: Knowledge Graph Attention Network for Recommendation," *Proc. the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.950-958, 2019.
- [2] Y. Lin, Z. Liu, M. Sun, Y. Liu and X. Zhu, "Learning Entity and Relation Embeddings for Knowledge Graph Completion," *Proc. Twenty-ninth AAAI conference on artificial intelligence*, pp.2181-2187, 2015.
- [3] 楽天株式会社, "楽天市場データ," 国立情報学研究所情報学研究データリポジトリ. (データセット), <https://doi.org/10.32130/idr.2.1>, 2014.