

購入アイテムの多様性を考慮した顧客分析モデルの提案

1X20C095-8 藤原大喜
指導教員 後藤正幸

1. 研究背景・目的

多くの EC サイトでは顧客との良好な関係構築や長期的な売上向上を目的として、購買履歴データを用いた様々な顧客分析が行われている。ビジネス施策立案のための購買履歴データの分析モデルも広く研究されており、多くのアプローチが提案されている。これらの研究の中で、顧客の“購買行動の多様性”が顧客の行動予測の予測精度向上に寄与するという報告がある [1]。また、顧客の購買行動の多様性自体をマーケティング活動のための分析の切り口としている研究も多い [2]。一方、商品、または商品カテゴリの購買頻度を用いたエントロピーやハフィンダール・ハーシュマン指数といった指標を用いて購買行動の多様性を評価する研究が従来提案されているが [1],[2]、アイテム間の類似性は考慮されていない。また、推薦システムで用いられる埋め込み表現モデルを用い、ユーザベクトルと購入アイテムベクトルの類似度の平均的な小ささを基準としたモデルも提案されている [3]。しかし、これらの従来手法は個々の購入アイテムが多様性に与える影響の違いを考慮せず、購買行動の多様性をスカラー値で表現している点で改善の余地がある。例えば、購入アイテムの多様性指標に与える影響が同程度であるユーザと、多様性指標に与える影響が強いアイテムとそうではないアイテムの双方を購入しているユーザが存在する可能性があるが、従来手法では前者と後者が購買行動の多様性の観点から同様の特徴をもつユーザであるとされてしまう。

そこで本研究では、推薦モデルの一種である Knowledge Graph Attention Network [4] (以下、KGAT) において個々の購入に対して付与される重みの分布を活用し、各顧客の購買行動の多様性を表す特徴量を算出することで、購入アイテムが多様性に与える影響を考慮した顧客分析を可能とする手法を提案する。最後に提案手法を実データに適用し、その有効性を示す。

2. 関連研究

2.1. 購買行動の多様性をスカラー値で表現するモデル

購買行動の多様性に関する顧客分析モデルでは、アイテムやそのカテゴリの購入頻度に基づくものが主であったが、購入されたアイテムの類似性を考慮しないことが多かった。そこで、Anderson ら [3] はユーザベクトルと購入アイテムベクトルの類似度に基づき購買行動の多様性をスカラー値で表現するモデルを提案した。しかし、このモデルでは、各ユーザの購買行動の多様性を表す指標がスカラー値で表現されており、購入されたアイテムごとの多様性に対する影響の違いを考慮した分析はなされていない。

2.2. Knowledge Graph Attention Network

KGAT [4] は購買履歴を活用して顧客の嗜好を捉える推薦モデルの一種であり、その有用性が示されている。KGAT は購買関係やアイテムとその補助情報の関係を表した協調知

識グラフを用いており、これは先頭エンティティ h 、リレーション r 、末尾エンティティ t からなる Triplet (h, r, t) の集合として定義される。ここでエンティティとはグラフにおけるノードを意味し、「ユーザ」、「アイテム」、「アイテムの補助情報」を示す。また、リレーションとはグラフにおけるエッジを意味し、「購買関係」と「アイテムとその補助情報との関係」を示す。KGAT における出力はユーザ u のアイテム i に対するマッチングスコア $\hat{y}(u, i)$ であり、式 (1) で定義される。

$$\hat{y}(u, i) = \mathbf{e}_u^\top \mathbf{e}_i \quad (1)$$

式 (1) において、 $\mathbf{e}_u, \mathbf{e}_i$ は、それぞれユーザ u 、アイテム i のベクトル表現を示す。KGAT の学習においてはまず、知識グラフ上のエンティティ、リレーションをベクトルとして表現する。その後、各エンティティのベクトル表現を更新する。このとき、該当エンティティを h とする Triplet における t のベクトル表現 \mathbf{e}_t を、式 (2) を用いて重み付けした和を \mathbf{e}_h に反映させることで更新する。

$$\pi(h, r, t) = \frac{\exp(\tilde{\pi}(h, r, t))}{\sum_{(h, r', t') \in \mathcal{N}_h} \exp(\tilde{\pi}(h, r', t'))} \quad (2)$$

$$\tilde{\pi}(h, r, t) = \mathbf{e}_t^\top \tanh(\mathbf{e}_h^* + \mathbf{e}_r^*) \quad (3)$$

式 (2) において、 \mathcal{N}_h は先頭エンティティ h に対して知識グラフ上で定義されている Triplet の集合を示す。また、式 (3) において、 $\mathbf{e}_h^*, \mathbf{e}_r^*, \mathbf{e}_t^*$ はそれぞれ、 $\mathbf{e}_h, \mathbf{e}_r, \mathbf{e}_t$ を同一空間上に射影したベクトル表現を示す。

3. 提案手法

3.1. 着想

学習済みの KGAT において、 h がユーザ、 r が購買、 t がアイテムである Triplet に対応する $\tilde{\pi}(h, r, t)$ の値が大きいことは、アイテム t を購入しているユーザ間で頻繁にみられる、すなわち類似した購買行動をユーザ h がしていることを示している。そのため、 $\tilde{\pi}(h, r, t)$ が大きい値に偏るユーザの購買行動は多様性が小さく、逆に小さい値に偏るユーザは多様なアイテムを購入する傾向にあると考えられる。さらに、それらの分布はユーザによって異なると考えられるため、 $\tilde{\pi}(h, r, t)$ の分布を用いて購買行動の多様性を分析可能であるといえる。

3.2. 提案手法の手続き

提案手法では学習済みの KGAT において、ユーザごとに全購入アイテムにおける $\tilde{\pi}(h, r, t)$ のヒストグラムを面積が 1 になるように正規化した上で作成し、そのビンの高さを並べたものを購買行動の多様性を表す特徴量とする。このとき、階級の設定は全ユーザで統一する。ユーザ個人に対し $\tilde{\pi}(h, r, t)$ の分布を捉えることで、そのユーザの購買行動の多様性を分析する。また、購買行動の多様性を表す特徴量を用いたクラ

スタリングにより、購買行動の多様性の全体的な大小に加え、 $\hat{\pi}(h, r, t)$ の分布傾向に関して類似したユーザをグループ分けする。

4. 実データ分析

本章では、提案手法の有効性を評価するために実データ分析を行う。まず、 $\hat{\pi}(h, r, t)$ の分布の平均が購入アイテムの多様性を表すことの妥当性を示す。次に、購買行動の多様性を表す特徴量を用いてクラスタリングをすることで、スカラー値で表現する指標より詳細に購買行動の多様性を分析できるようになることを示す。

4.1. 分析条件

本実験では、オンライン音楽サービスにおける鑑賞履歴データ及び楽曲情報データである Last-FM データセット [5] を用いる。ユーザ数は 23,566、曲数は 48,123、鑑賞履歴数は 3,034,796 である。エンティティのベクトルの次元数は 64 とした。購買行動の多様性を表す特徴量を用いたクラスタリングには k-means 法を用い、クラスタ数は 10 とした。 $\hat{\pi}(h, r, t)$ のヒストグラムに関しては、階級の数 を 20、表示する階級の上限、下限をそれぞれ全購買における $\hat{\pi}(h, r, t)$ の最大値、最小値である 15.08、-0.07280 として作成した。

4.2. 分析結果

$\hat{\pi}(h, r, t)$ の分布の形がおおよそ釣鐘型で共通しており、分散がほとんど等しく、平均に違いが見られる 2 人のユーザ (以下、ユーザ A、ユーザ B) に対して、ユーザベクトルとそれぞれのユーザに対する $\hat{y}(u, i)$ 上位 100 個のアイテムベクトルを t-SNE を用いて 2 次元に可視化した結果を図 1 に示す。平均はユーザ A が 7.37、ユーザ B が 5.48 であった。分散はユーザ A が 0.84、ユーザ B が 0.87 であった。

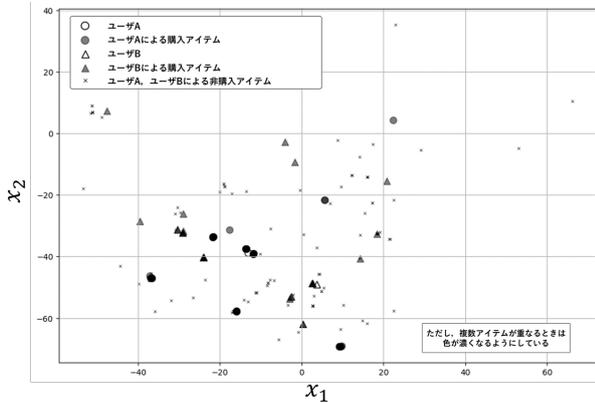


図 1: ユーザと $\hat{y}(u, i)$ のランキング上位 100 個のアイテム

図 1 より、 $\hat{\pi}(h, r, t)$ の平均が比較的大きいユーザ A に関しては、鑑賞した曲のベクトルがユーザベクトルの近くにまとまるように分布していることがわかる。一方、 $\hat{\pi}(h, r, t)$ の平均が比較的小さいユーザ B に関しては、鑑賞した曲のベクトルがユーザベクトルから離れて広範囲にばらつくように分布していることがわかる。これらのことから、ユーザごとの $\hat{\pi}(h, r, t)$ の平均の大きさは鑑賞する曲の多様性の小ささ、つまり鑑賞行動の多様性の小ささを示すことがわかる。

次に、購買行動の多様性を表す特徴量を用いたクラスタリングの結果として、各ユーザの $\hat{\pi}(h, r, t)$ の分布の平均と分散の各クラスタにおける平均と標準偏差を表 1 に示す。

表 1: 購買行動の多様性を表す特徴量を用いたクラスタリングの結果

クラスタ	1	2	3	4	5
平均	5.18(± 0.62)	6.07(± 0.17)	6.56(± 0.19)	6.78(± 0.21)	7.20(± 0.16)
分散	0.94(± 0.20)	0.80(± 0.37)	0.62(± 0.23)	0.58(± 0.27)	0.29(± 0.16)
クラスタ	6	7	8	9	10
平均	7.22(± 0.28)	7.55(± 0.19)	7.75(± 0.24)	7.84(± 0.28)	8.45(± 0.21)
分散	0.90(± 0.35)	0.46(± 0.20)	0.72(± 0.21)	0.40(± 0.23)	0.58(± 0.47)

表 1 より、 $\hat{\pi}(h, r, t)$ の平均と分散にはクラスタごとに違いがみられる。クラスタリングにより、購買行動の多様性の観点からユーザをグループ分けできている。

5. 考察

図 1 より、ユーザ A は鑑賞した曲の多様性が比較的小さいことから、鑑賞行動の多様性が小さいと考えられる。一方、ユーザ B は鑑賞した曲の多様性が比較的大きいことから、鑑賞行動の多様性が大きいと考えられる。このように、 $\hat{\pi}(h, r, t)$ の分布の平均を用いることで、ユーザごとに購買行動の多様性の違いを分析することが可能になる。また、表 1 より、クラスタ 5 とクラスタ 6 では $\hat{\pi}(h, r, t)$ の平均が近い値であるが、分散に差異がみられる。すなわち、ユーザの購買行動の多様性の大きさだけでなく、その傾向の違いも考慮できているといえる。ユーザごとにスカラー値として購買行動の多様性を表現する従来手法では、このような分析は困難であり、提案手法の独自性があるポイントといえる。KGAT ではアイテムの補助情報も学習に用いているため、各アイテムの補助情報を用いて顧客の嗜好をさらに深掘りすることも可能となる。

6. まとめと今後の課題

本研究では、KGAT において各購買に付与される重みを用いることで、購入アイテムが多様性に与える影響を考慮した顧客分析を可能とする手法を提案した。また、提案手法を実データに適用することで、提案手法の妥当性と分析例を示した。今後の課題としては、購買行動の多様性を表す特徴量を用いたクラスタリングにおける適切なクラスタ数の検証などが挙げられる。

参考文献

- [1] 新美潤一郎, 星野崇宏. 顧客行動の多様性変数を利用した購買行動の予測 Deep Learning を応用した実店舗・Web・モバイルの多面的な分析. 人工知能学会論文誌, Vol. 32, No. 2, pp. B-G63-1, 2017.
- [2] 森田裕之, 白井康之, 中元政一. エントロピーを応用したフラッシュマーケティングにおける多様性を考慮した分析. 経営情報学会 全国研究発表大会要旨集, Vol. 2013, No. 0, pp. 329-332, 2013.
- [3] Ashton Anderson, Lucas Maystre, Ian Anderson, Rishabh Mehrotra, and Mounia Lalmas. Algorithmic Effects on the Diversity of Consumption on Spotify. In *Proceedings of the Web Conference 2020*, pp. 2155-2165, 2020.
- [4] 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 and Data Mining*, pp. 950-958, 2019.
- [5] T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere. The million song dataset. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, 2011.