

ユーザの関心度を考慮した多変量ホークス過程に基づく音楽推薦手法 Music Recommendation Method Based on Multivariate Hawkes Process Considering User Interest Levels

山田 航平 原川 良介 岩橋 政宏
Kohei Yamada Ryosuke Harakawa Masahiro Iwahashi

概要

ユーザの音楽聴取履歴を基にして、次に聴取する音楽を推薦する手法が種々提案されている。近年の多変量ホークス過程に基づく推薦手法は、強度関数によってユーザの関心の変化を把握可能とし、推薦を実現している。ここで、ユーザが聴取した音楽は全て関心がある音楽であると仮定されているが、この仮定は必ずしも成立しないという報告が存在する。本文では、各音楽の演奏時間のうちユーザが実際に聴取した時間に基づき関心度を算出し、推薦結果に反映可能とする新たな手法を提案する。提案手法では、音楽ごとに算出された関心度に応じて、多変量ホークス過程の強度関数の値を変化させることで、ユーザの関心が高い音楽をより上位に推薦可能とする。適応的な強度関数を導出した点が、全ての音楽に対する関心度を固定していた従来の推薦手法と比較して新規である。広く用いられている公開データセットである“lastfm-dataset-1K”に対する実験によって、提案手法が各ユーザに適した音楽を推薦可能とすることを示した。

1 序論

Last.fm¹⁾などの音楽ストリーミングサービスの普及に伴い、ユーザの聴取履歴を基にして次に聴取する音楽を推薦する手法が種々提案されている。しかしながら、ユーザの関心は時間とともに変化するため、ユーザが望む音楽を推薦することは依然として困難な課題である [1,2]。

従来手法として、ユーザが聴取した音楽の内容を考慮した手法 [3,4] が挙げられる。これらの手法は特定期間におけるユーザの聴取履歴を解析し、ユーザの長期的な関心に基づいて推薦を実行している。しかしながら、ユーザの関心は時間の経過とともに変化する場合があるため、短期的な関心を加味することは音楽を推薦する上で重要である。

文献 [1,5,6] では、ユーザが過去に聴取した音楽の内容のみならず、聴取履歴の順序、すなわちユーザが最近どのような音楽を聴取したのかを加味した手法が提案されている。文献 [1] では、多変量ホークス過程の強度関数によって、ユーザの長期的な関心だけでなく、短期的な関心を捉えられる。さらに、

attention 機構により、各ユーザが長期的・短期的関心のどちらを優先するのかを適応的に解析可能としている。ここで、文献 [1] では、聴取された音楽全てをユーザが関心を持った音楽であると仮定している。

一方で、音楽ストリーミングサービスの利用者の多くが、聴取開始直後に音楽をスキップした経験があると報告されている²⁾。より具体的には、ユーザが音楽を素早く評価し、関心が低ければ短時間の聴取で切り替える傾向があることが示されている。したがって、ユーザが聴取した音楽それぞれに対する関心度を加味することで、従来手法の更なる高精度化が期待できる。

本文では、ユーザごとに各音楽に対する関心度を算出し、推薦結果に反映可能とする新たな手法を提案する。提案手法では、各音楽の演奏時間のうちユーザが実際に聴取した時間に基づき関心度を算出する。さらに、算出された関心度を強度関数の短期的な関心を捉える項に適応させる。これにより、文献 [1] では音楽の埋め込みベクトルのみが強度関数の変化量を決定していたが、提案手法では各音楽に対する関心度を加味して強度関数の変化量を算出可能となる。本研究では、音楽推薦の研究に広く用いられている“lastfm-dataset-1K”データセット³⁾ [7,8] を用いて、提案手法によって各ユーザに適した音楽が推薦可能となることを示した。

2 提案手法

2.1 記号

表 1 に本文で使用する主要な記号を示す。聴取履歴内の各音楽 $i \in \mathcal{I}$ は対応する埋め込みベクトル $v_i \in \mathbb{R}^D$ に変換される。ここで、 \mathcal{I} と D は音楽の集合と次元数である。 v_i を $|\mathcal{I}|$ 個の音楽に対応する行列にすることで、音楽の埋め込み行列 $V \in \mathbb{R}^{D \times |\mathcal{I}|}$ と表す。具体的には、 $v_i = V\bar{v}_i$ となる。ここで、 $\bar{v}_i \in \mathbb{R}^{|\mathcal{I}|}$ は、音楽を一意に識別するために 1 つの次元に 1 があり、他の全ての次元は 0 となる one-hot ベクトルである。同様に、ユーザ u の埋め込みベクトル $v_u \in \mathbb{R}^D$ は、埋め込み層でユーザの埋め込み行列 $U \in \mathbb{R}^{D \times |\mathcal{U}|}$ を参照することで取得できる。

2.2 ユーザの関心度

本研究では、ユーザの関心度 M_i を、各音楽の演奏時間のうちユーザが実際に聴取した時間に基づき定

長岡技術科学大学 大学院工学研究科
本研究の一部は JSPS 科研費 JP23K28193 およびセコム科学技術振興財団の助成を受けたものです。

1) <http://www.last.fm>

2) <https://musicmachinery.com/2014/05/02/the-skip/>

3) <https://github.com/eifuentes/lastfm-dataset-1K?tab=readme-ov-file>

表 1 表記と記号

記号	説明
\mathcal{U}, \mathcal{I}	ユーザの集合と音楽の集合
$u \in \mathcal{U}, i \in \mathcal{I}$	ある特定のユーザとある特定の音楽
S_u	ユーザ u の聴取履歴の順序
$\mathcal{H}_{u,t} \subseteq S_u$	ユーザ u の時刻 t より過去の聴取履歴の順序
$U \in \mathbb{R}^{D \times \mathcal{U} }$	ユーザの埋め込み行列 ($ \cdot $ は集合の要素数)
$V \in \mathbb{R}^{D \times \mathcal{I} }$	音楽の埋め込み行列
$v \in \mathbb{R}^D$	各ユーザまたは各音楽の D 次元埋め込みベクトル
M_i	音楽 i に対するユーザの関心度
γ	関心度に対する重み
$\mu_{u,i}$	ユーザ u に対する音楽 i の 強度関数の定数
$\alpha_{h,i}$	過去に聴取した音楽 h が 音楽 i の強度関数を変化させる量
$\kappa(\cdot)$	強度関数の時間的減衰 を計算するカーネル関数
δ_u	ユーザ u の カーネル関数の時間的減衰率
β_u^l, β_u^s	ユーザ u の長期的, 短期的な 関心の重み
$\tilde{\lambda}_{qu}(t)$	時刻 t におけるユーザ u の 音楽 i に対する関心の予測値

義する. 具体的には, 計算式は次のようになる.

$$M_i = \frac{\text{音楽 } i \text{ の聴取時間}}{\text{音楽 } i \text{ の演奏時間}} \quad (1)$$

ここで, 式 (1) の分子は, ユーザが音楽を聴取した時刻と次の音楽を聴取した時刻の差を表す. M_i の最大値と最小値はそれぞれ 1 と 0 となる. M_i が大きいほど, ユーザがその音楽を長く聴いたことを示し, その音楽への関心が高いことを意味する. M_i が小さいほど, ユーザがその音楽を早くスキップしたことを示し, その音楽への関心が低かったことを意味する.

2.3 関心度を考慮した多変量ホークス過程

提案手法には, ユーザの聴取履歴の順序 S_u , 音楽の埋め込み行列 V , ユーザの埋め込み行列 U , ユーザの関心度 M_i , および音楽が聴取された時刻 t_k を入力する. 出力はユーザが次に聴取する確率が高い音楽を順番に出力する. 提案手法の新規性は, 音楽ごとに算出された関心度に応じて, 多変量ホークス過程の強度関数の短期的な関心を表す項を変化させることで, ユーザの関心が高い音楽をより上位に推薦可能とすることである.

多変量ホークス過程を採用して, 聴取履歴の順序に基づいて強度関数を算出し, 次に聴取する音楽

(ターゲット音楽) を推薦する. 条件付き強度関数は, ユーザが音楽を聴取する強度である. 具体的には, ユーザ $u \in \mathcal{U}$ および u の過去の聴取履歴の順序が与えられた場合, 時刻 t にターゲット音楽 $i \in \mathcal{I}$ が聴取されるとい条件付き強度関数は, 次式によって定義される.

$$\tilde{\lambda}_{qu}(t) = \beta_u^l \mu_{u,i} + \beta_u^s \sum_{h \in \mathcal{H}_{u,t}} \alpha_{h,i} (\exp(\gamma M_i) - 1) \kappa_u(t - t_h) \quad (2)$$

式 (2) は, ユーザ u の長期的な関心と短期的な関心を表現す. 右辺の第 1 項が長期的な関心を表し, 第 2 項が短期的な関心を表す. 具体的には, ユーザが音楽を聴取した際, 全ての音楽の強度関数が第 2 項によって最初に増大または抑制され, その後, 以前聴取した音楽の影響が 0 に向かって減衰するにつれて第 1 項に近づく.

式 (2) の $\mu_{u,i} (= v_u^T v_i)$ は, ユーザ u に対する音楽 i の強度関数の定数であり, u の埋め込みベクトルと i の埋め込みベクトルの一致度に依存する. $\mathcal{H}_{u,t} \subseteq S_u$ は, 時刻 t 以前のユーザ u の過去の聴取履歴の順序を表し, 特に計算コストを削減するために固定長の部分的な聴取履歴の順序として S_u から切り取られる. 本研究では, $\mathcal{H}_{u,t}$ は直近の聴取履歴の順序から 3 つ前の履歴までを使用する ($|\mathcal{H}_{u,t}| = 3$). $\alpha_{h,i} (= v_h^T v_i)$ は, 過去に聴取した音楽 h が音楽 i の強度関数を変化させる量であり, h および i の埋め込みベクトルに依存する. v_h は音楽 h の埋め込みベクトルである. β_u^l, β_u^s は attention 機構によってユーザの長期および短期の関心の重みを決定する [9, 10].

$\kappa_u(\cdot)$ は次式で定義される指数カーネル関数であり, 時間とともに強度関数が指数関数的に減衰することを表す.

$$\kappa_u(t - t_h) = \exp(-\delta_u(t - t_h))$$

ここで, $\delta_u \geq 0$ はユーザ u 毎の指数カーネル関数の時間的減衰率を表し, 各ユーザの関心が異なる速度で減衰する可能性があるため, ユーザ依存のパラメータである.

式 (2) における $(\exp(\gamma M_i) - 1)$ は, M_i によって指数カーネル関数の変化量を変化させるためのものである. これにより, 文献 [1] と比較して, 提案手法における多変量ホークス過程は, ユーザの関心度に応じて条件付き強度関数を変化させる量を変化させ, 次にユーザが高い関心を持つ音楽を推薦可能とする.

次に, 推薦する音楽の決定方法を説明する. ユーザ u が時刻 t にターゲット音楽 i に関心を持つ確率を次のように定義する.

$$p_{qu}(t) = \frac{\exp(\tilde{\lambda}_{qu}(t))}{\sum_{k \in \mathcal{I}} \exp(\tilde{\lambda}_{ku}(t))} \quad (3)$$

ターゲット音楽 $i \in \mathcal{I}$ に対して, 式 (3) は音楽の集合

I 全体に対する条件付き分布 $p_{iu}(t)$ を定義する。最後に、2つの音楽 i と i' のランキングスコアに従って、以下のように推薦を実行する。

$$i >_{u,t} i' : \Leftrightarrow p_{iu}(t) > p_{i'u}(t).$$

ここで、 $>_{u,t}$ は時刻 t でのユーザ u に対する推薦候補の音楽の順位を表す。

学習過程では、式 (3) は学習データ内の全てのユーザの聴取履歴の順序に対して最大化される。ネガティブサンプリング [11] を用いることで、式 (3) を効率的に計算できる。次に、対数確率を次のように定義する。

$$\log p_{iu}(t) \propto \log \sigma(\tilde{\lambda}_{iu}(t)) + n \cdot E_{r \sim p_r} [\log \sigma(-\tilde{\lambda}_{r'u}(t))] \quad (4)$$

ここで、 $\sigma(x)$ はシグモイド関数、 n はネガティブサンプルの数、 r' は一様分布に基づいて音楽の集合からサンプリングされた音楽である。また、ネガティブサンプルの数 n は音楽の集合の要素数 $|I|$ よりも十分小さく、学習時間は音楽の集合の要素数 $|I|$ に依存しない。従来の最適化方法（確率的勾配降下法など）を使用して、式 (4) で定義された目的関数を最適化できる。

3 実験

本章では、提案手法の有効性を確認するため、以下を検証する。

- 提案手法は精度よく音楽を推薦可能か。
- 提案手法はどのようなユーザに特に有効か。

以降では、3.1 節でデータセットとハイパーパラメータについて述べ、3.2 節で評価指標と比較手法を説明し、3.3 節で評価結果を示す。

3.1 データセットとハイパーパラメータ

Lastfm-dataset-1K³⁾ は Last.fm¹⁾ の音楽を含んでいる。具体的には、2006 年から 2009 年の 60 人のユーザによる 1,048,576 件の聴取履歴が含まれている。

このデータセットを学習データとテストデータに重複しないように分割した。学習データは、総ユーザの 80% の聴取履歴の順序と、残り 20% のユーザの聴取履歴の順序の前半から構成され、後半の聴取履歴の順序はテストデータとして用いた。具体的には、各ユーザ u の聴取履歴の順序 $S_u : \{(u, i(1), t_1), (u, i(2), t_2), \dots, (u, i(n_u), t_{n_u})\}$ に対して、 k 番目の学習データは $\{(u, i(1), t_1), (u, i(2), t_2), \dots, (u, i(k), t_k)\}$ と表される。

次に、提案手法のハイパーパラメータの設定を説明する。バッチサイズ、エポック数、ネガティブサンプル数をそれぞれ 512, 50, 5 に設定した。ネットワークの次元サイズは $D=16, 32, 64$ と定義した。提案手法のパラメータの最適化には、学習率 0.001 の Adam [12] を用いた。

3.2 評価指標と比較手法

本実験では、提案手法と後述の比較手法によって生成された上位 k 件の推薦結果の妥当性を評価した。評価にはまず、次式の Recall を使用した。

$$\text{Recall}@k = \frac{\#(\text{hit})}{\#(\text{testdata})}$$

ここで、 k は推薦結果の音楽の数、 $\#(\text{hit})$ はヒットの数、 $\#(\text{testdata})$ はテストデータの総数である。ここで、ヒットとは、ターゲット音楽が推薦結果に含まれていることを意味する。例えば、あるテストデータの正解が i_t で、 i_t が推薦リストに含まれている場合にヒットと言う。

さらに、次式の MRR を評価に使用した。

$$\text{MRR}@k = \frac{\sum_n (1/\text{rank}_n)}{\#(\text{testdata})}$$

ここで、 rank_n は、 n 番目のテストデータに対する推薦結果内のターゲット音楽のランクである。MRR は、推薦リストにおけるターゲット音楽のランクの逆数の平均を示す。

実験では、提案手法と次の手法を比較した。

- SHAN [5] は、ユーザの長期および短期の attention を取得するために 2 層の階層的 attention ネットワークを構築し、ユーザの特徴を表現する。
- RDR [13] は、skip-gram 手法 [11] を使用して全音楽の分散表現を算出する。
- MHPE-a [1] は、多変量ホークス過程と attention 機構を使用して、ユーザの長期および短期の行動パターンを適応的に表現できる。

RDR [13] と比較することで、提案手法において聴取履歴の順序を加味することの有効性を確認することができる。また、SHAN [5] と MHPE-a [1] と比較することで、提案手法におけるユーザの関心度を考慮することの有効性を明らかにできる。

3.3 結果

3.3.1 比較手法との性能比較

比較結果を表 2 に示す。提案手法が Recall と MRR の双方で比較手法よりも高い精度を達成している。これは、ユーザの聴取履歴の順序だけでなく、関心度を用いることで、ユーザの望む音楽をより上位に推薦可能となったことを示している。具体的には、多変量ホークス過程により、ユーザの聴取履歴の順序から埋め込みベクトルを学習できるため、音楽の内容を考慮した手法 (RDR) よりも優れた結果を獲得できた。さらに、提案手法はユーザの聴取履歴の順序に加え、ユーザの短期的な関心をより正確に捉えられたため、他の聴取履歴の順序を考慮した手法 (SHAN, MHPE-a) よりも優れた結果を得られた。

表2 lastfm-dataset-1Kにおける性能比較

		Recall@k		MRR@k	
		@10↑	@20↑	@10↑	@20↑
D=16	SHAN [5]	0.0491	0.0892	0.00975	0.0113
	RDR [13]	0.0678	0.114	0.0145	0.0171
	MHPE-a [1]	0.0921	0.150	0.0260	0.0310
	提案手法	0.115	0.189	0.0398	0.0420
D=32	SHAN [5]	0.0597	0.101	0.0148	0.0160
	RDR [13]	0.0885	0.134	0.0175	0.0209
	MHPE-a [1]	0.130	0.178	0.0339	0.0375
	提案手法	0.151	0.217	0.0459	0.0477
D=64	SHAN [5]	0.0686	0.122	0.0196	0.0225
	RDR [13]	0.106	0.147	0.0201	0.0247
	MHPE-a [1]	0.150	0.200	0.0372	0.0401
	提案手法	0.182	0.239	0.0491	0.0524

表3 ユーザタイプ毎の性能比較

	Recall@k		MRR@k	
	@10	@20	@10	@20
全ユーザ	0.182	0.239	0.0491	0.0524
ユーザタイプ(1)	0.236	0.278	0.0589	0.0645
ユーザタイプ(2)	0.176	0.234	0.0378	0.0458

3.3.2 ユーザタイプによる性能比較

本項では、提案手法の有効範囲を検証する。まず、各ユーザの関心度の分散の中央値を計算した。さらに、分散が中央値より大きいユーザを、最後まで聴取する音楽とスキップする音楽が混在しているユーザとして、ユーザタイプ(1)と定義した。一方、中央値より小さいユーザを、ほとんどの音楽を最後まで聴取するユーザとして、ユーザタイプ(2)と定義した。表3は、提案手法を $D=64$ で学習させた場合のRecallおよびMRRを示している。結果より、提案手法は特にユーザタイプ(1)に対して有効であることが明らかとなった。したがって、音楽毎でユーザの関心度に明確な違いが存在した場合、提案手法が正確にそれを反映できたことが示唆された。

4 まとめ

提案手法は、各音楽の実際に聴取した時間に基づき関心度を算出し、多変量ホークス過程の強度関数の値を変化させることで、ユーザの関心が高い音楽をより上位に推薦可能とした。広く用いられている公開データセットである“lastfm-dataset-1K”に対する実験によって、提案手法が各ユーザに適した音楽

を推薦可能であることを示した。

参考文献

- [1] D. Wang, X. Zhang, Z. Xiang, D. Yu, G. Xu, and S. Deng, “Sequential recommendation based on multivariate hawkes process embedding with attention,” *IEEE Trans. Cybernetics*, vol. 52, no. 11, pp. 11 893–11 905, 2022.
- [2] K. Ohtomo, R. Harakawa, M. Iisaka, and M. Iwahashi, “Ambi-lstm: Adaptive multi-modal bi-lstm for sequential recommendation,” *IEEE Access*, vol. 12, pp. 12 720–12 733, 2024.
- [3] S. Kangas, “Collaborative filtering and recommendation systems,” *VTT Information Technology*, pp. 18–20, 2002.
- [4] A. Van den Oord, S. Dieleman, and B. Schrauwen, “Deep content-based music recommendation,” in *Proc. Advances in Neural Information Processing Systems*, 2013, pp. 1–9.
- [5] H. Ying, F. Zhuang, F. Zhang, Y. Liu, G. Xu, X. Xie, H. Xiong, and J. Wu, “Sequential recommender system based on hierarchical attention networks,” in *Proc. Int. Joint Conf. Artificial Intelligence*, 2018, pp. 3926–3932.
- [6] W.-C. Kang and J. McAuley, “Self-attentive sequential recommendation,” in *Proc. IEEE Int. Conf. Data Mining*, 2018, pp. 197–206.
- [7] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos, “A unified framework for providing recommendations in social tagging systems based on ternary semantic analysis,” *IEEE Trans. Knowledge and Data Engineering*, vol. 22, no. 2, pp. 179–192, 2010.
- [8] H. Weng, J. Chen, D. Wang, X. Zhang, and D. Yu, “Graph-based attentive sequential model with metadata for music recommendation,” *IEEE Access*, vol. 10, pp. 108 226–108 240, 2022.
- [9] D. Wang, X. Zhang, D. Yu, G. Xu, and S. Deng, “Came: Content- and context-aware music embedding for recommendation,” *IEEE Trans. Neural Networks and Learning Systems*, vol. 32, no. 3, pp. 1375–1388, 2021.
- [10] J. Han, L. Zheng, Y. Xu, B. Zhang, F. Zhuang, P. S. Yu, and W. Zuo, “Adaptive deep modeling of users and items using side information for recommendation,” *IEEE Trans. Neural Networks and Learning Systems*, vol. 31, no. 3, pp. 737–748, 2020.
- [11] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in *Proc. Advances in Neural Information Processing Systems*, 2013, pp. 1–9.
- [12] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv:1412.6980*, 2014.
- [13] D. Wang, S. Deng, S. Liu, and G. Xu, “Improving music recommendation using distributed representation,” in *Proc. Int. Conf. Companion on World Wide Web*, 2016, pp. 125–126.