

# 楽曲特徴に基づく音楽トレンド予測およびアーティスト発掘

河村 和紀 宇田川 拓麻 吉田 将大 舘野 啓

ソニーグループ株式会社

## 1 はじめに

近年、音楽アーティスト発掘はオーディションやライブ会場などのオフラインではなく、SNS や音楽プラットフォームなどオンライン上でおこなわれることが多くなっている [1]. このようなインターネット上での発掘は、発掘対象のアーティストが膨大になるため、レコード会社の A&R (アーティストの発掘を担う職種) は、膨大な数のアーティストの人気度を自動で追跡し、人気度が急上昇するタイミング、すなわち、アーティストの流行り始めを検知するツールを導入している。

しかしながら、このようなツールは既に発掘の現場に広く普及しているため、より早く有望なアーティストを発掘するためには、流行り始めるよりも前に今後の流行りを予測する必要がある。そこで、先行研究では楽曲の人気度の時系列変化を入力とし LSTM を用いて将来の人気度を予測する手法が提案されている [2]. しかし、このように楽曲単体の人気度変化を追跡する手法では、新人アーティストのように人気度が低くかつ変化がほとんど見られないアーティストの楽曲の時系列変化を予測するのは困難である。

そこで本研究では、アーティスト側の変化ではなくより大局的な変化である音楽を聴く消費者側の音楽的なトレンドの変化を予測することに焦点をあて、予測された音楽トレンドをもとに、将来人気度が上昇すると予想されるアーティストを先んじて発掘する手法を提案する。まず、各楽曲をテンポ、キーなどの音楽特徴の集合で表し、各期間ごと各音楽特徴ごとに人気があった楽曲群を度数分布で表現する。そして、ニューラルネットワーク (NN) を用いてその度数分布を時系列予測する。さらに、現在の分布と将来の分布を比較し、顕著に分布数が増加する音楽特徴量区間をその音楽特徴の音楽トレンドとする。最後に、音楽特徴がこの音楽トレンドに合致するような楽曲を探索し、その楽曲のアーティストを抽出する。

## 2 提案手法

### 2.1 問題設定

はじめに、本論文における音楽トレンドを定義する。まず、音楽配信サービスのヒットチャートや消費者の再生履歴をもとに、各期間内における人気の楽曲を抽出し、それらの楽曲のテンポやキーなどの音楽特徴の分

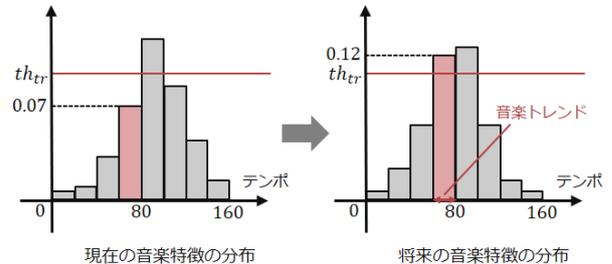


図 1: 音楽トレンドの定義

布を考える。続いて、図 1 のように注目する期間と一時刻前の期間の分布を比較し、前の期間では閾値  $th_{tr}$  より分布数が低く、対象の期間では閾値より分布数が高くなるような特徴量の区間を音楽トレンドとよぶ。

音楽トレンドは、注目する期間の分布が求まり、閾値が与えられれば自動的に決まる。したがって、音楽トレンドを予測するという事は注目する期間の音楽特徴分布を予測することに他ならない。各期間  $t$  の各音楽特徴  $f \in \mathcal{F}$  の分布は、正規化した度数分布

$$p_f^t \left( f_{\min} + (n_b - 1) \frac{f_{\max} - f_{\min}}{N_b} \leq X < f_{\min} + n_b \frac{f_{\max} - f_{\min}}{N_b} \right) = \frac{1}{N_d} N_{[f_{\min} + (n_b - 1) \frac{f_{\max} - f_{\min}}{N_b}, f_{\min} + n_b \frac{f_{\max} - f_{\min}}{N_b}]}$$

で表される。ここで、 $N_d$  は期間  $t$  の楽曲のうち人気のある楽曲の総数、 $N_{[m,n]}$  は、そのうち、楽曲特徴量  $x$  の値が  $m \leq x < n$  となる楽曲数、 $f_{\min}, f_{\max}$  はそれぞれ楽曲特徴量の最小値と最大値、 $N_b$  は階級数である。ここで、 $N_b$  は各特徴ごとに設定され、全期間通して同じ値が用いられる。このとき、トレンド予測の目的は、 $[1, T]$  の期間の分布  $p_{f \in \mathcal{F}}^{t \in [1, T]}(X)$  を用いて、 $[T+1, T+T']$  の分布  $p_{f \in \mathcal{F}}^{t \in [T+1, T+T']}(X)$  を予測することである。また、アーティスト発掘の目的は、各期間ごとの楽曲総再生回数などから計算される各アーティスト  $a \in \mathcal{A}$  の人気度を  $po_a^t$  としたとき、ある閾値  $th_{po}$  を持って定められるアーティストの集合  $\mathcal{A}_p = \{a \in \mathcal{A} \mid po_a^T < th_{po}, th_{po} \leq po_a^{T+T'}\}$  を求めることである。

### 2.2 トレンド予測およびアーティスト発掘

$p_{f \in \mathcal{F}}^{t \in [1, T]}(X)$  が与えられたとき、図 2 のように NN を用いて  $p_{f \in \mathcal{F}}^{t \in [T+1, T+T']}(X)$  を予測する。

$$(y_{T+1}, y_{T+2}, \dots, y_{T+T'}) = \text{NN}(y_1, y_2, \dots, y_T),$$

Forecasting music trends and finding new artists based on music features: Kazuki Kawamura, Takuma Udagawa, Masahiro Yoshida, Kei Tateno (Sony Group Corporation)  
Kazuki.Kawamura@sony.com

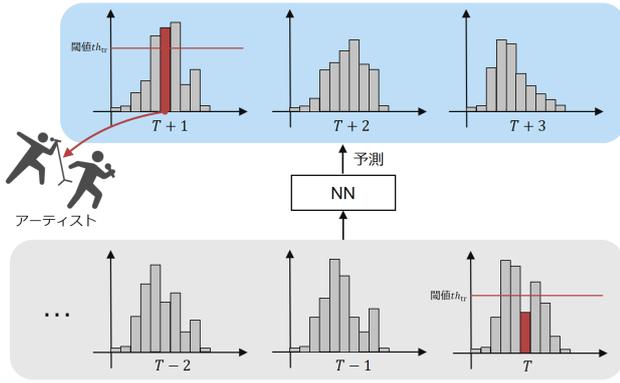


図 2: トレンド予測およびアーティスト発掘手法

$$\text{where } y_t = \bigcup_{f=f_1}^{f_{|\mathcal{F}|}} \bigcup_{n_b=1}^{N_b} p_f^t \left( f_{\min} + (n_b - 1) \frac{f_{\max} - f_{\min}}{N_b} \leq X < f_{\min} + n_b \frac{f_{\max} - f_{\min}}{N_b} \right).$$

NNとしてはMLP, LSTM, Transformerを用いる。注目している期間  $T + T'$  の分布が予測できたら、音楽トレンドは各楽曲特徴ごとに  $\mathcal{T}_f = \{X \mid p^T(X) < th_{tr}, th_{tr} \leq p^{T+T'}(X)\}$  として求める。

音楽トレンド  $\mathcal{T}_f$  が求まったら、時刻  $T + T'$  において、楽曲特徴が  $th_{so}$  個以上合致する楽曲  $s \in \mathcal{S}$ , すなわち、 $\mathcal{S}_d = \{s \in \mathcal{S} \mid |f \in \mathcal{F} : v_{sf} \in \mathcal{T}_f| \geq th_{so}\}$  を求め、これらの楽曲のアーティスト  $\mathcal{A}_d$  を抽出する。ここで、 $v_{sf}$  は、各楽曲  $s$  の楽曲特徴  $f$  の特徴量である。最後に、提案手法で用いる代表的な記法を表 1 にまとめる。

### 3 評価実験

#### 3.1 実験条件

音楽配信サービス Last.fm ストリーミング履歴のデータセット [3] を用いて、トレンド予測とアーティスト発掘手法を評価する。このデータセットは、992 人のユーザの 2006 年 8 月～2009 年 6 月の履歴で、各レコードにはユーザ ID, タイムスタンプ, アーティスト名, 楽曲名が含まれる。特定の期間に人気のある楽曲は特定の期間での再生回数に基づいて算出した。このデータセット自体に楽曲の音楽特徴は含まれないため、大規模楽曲データベースである Million Song Dataset [4] を照らし音楽特徴を取得した。データセットを 1 か月ごとに区切り、トレンド予測の実験では  $T' = 3$ , アーティストの発掘では  $T' = 1$  と設定した。

トレンド予測の実験では、データセットを学習・検証・テスト用に分割し、テスト期間の予測性能を対数尤度で評価する。具体的には、予測した分布  $\hat{p}^t(X)$  によって得られる十分量  $N_s$  のサンプルの実際の分布における対数尤度

$$\frac{1}{N_s} \sum_{i=1}^{N_s} \log p^t(x_i), \quad x_i \sim \hat{p}^t(X)$$

表 1: 本論文で使う代表的な記号

記号	説明
$s$	楽曲
$f$	楽曲特徴
$a$	アーティスト
$f_{\min}, f_{\max}$	楽曲特徴量の最大値, 最小値
$v_{sf}$	楽曲 $s$ の楽曲特徴 $f$ の特徴量
$\mathcal{T}_f$	各特徴ごとの音楽トレンド
$\mathcal{S}_d$	音楽トレンドに合致する楽曲集合
$\mathcal{A}_p$	本来発掘すべきアーティスト集合
$\mathcal{A}_d$	実際発掘したアーティスト集合
$th_{po}$	アーティストの人気度の閾値
$th_{tr}$	各楽曲特徴のトレンドの閾値
$th_{so}$	一致楽曲数の閾値

表 2: トレンド予測性能

モデル	1month	2month	3month
last	1	0.9984	0.9989
MLP	1.018	1.017	1.018
LSTM	0.9982	0.9982	0.9983
Transformer	0.9975	0.9971	0.9971

last の 1 か月後の予測性能で正規化 (低いほど性能が良い)

表 3: アーティスト発掘性能

モデル	Recall	Precision
last	0.0116	0.0498
Transformer	0.0135	0.0562

で評価する。このとき、5 時刻分のバックテストをおこない、各テストは 10 回試行したのち平均値を評価値とする。ベースラインは学習データの最後の時刻の分布  $p^T$  を予測分布として扱ったときの対数尤度とする。

アーティスト発掘の実験では、時刻  $T$  の時点で時刻  $T+1$  のトレンドに合致するアーティストを抽出し、実際には抽出対象であるアーティスト集合  $\mathcal{A}_d = \{a \mid po_a^T < th_{po}, th_{po} \leq po_a^{T+1}\}$  と比較し、発掘性能を Recall, Precision で評価する。ベースラインとしては、学習データの最後の時刻  $T$  と一つ前の時刻  $T-1$  との分布を比較してアーティストを発掘したときの性能を用いる。

トレンド予測の実験結果を表 2 に示す。この表から、1, 2, 3 か月先すべての予測において Transformer が最も高い性能で、いずれの月もベースラインの予想性能を上回っていることが分かる。また、表 3 にアーティスト発掘の結果を示す。これより、Transformer を用いた手法がベースラインよりも高い性能を示すことが分かる。

### 参考文献

- [1] S. Storstadmo *et al*, “In which ways has the digitalization changed the A&R profession,” *Master’s Thesis, University of Agder*, 2017.
- [2] Z. Wang *et al*, “Music Trend Forecast Based on LSTM,” *ICCIA*, 30–35, 2019.
- [3] O. Celma *et al*, “Music Recommendation and Discovery in the Long Tail,” *Springer*, 2010.
- [4] T. Bertin-Mahieux *et al*, “The Million Song Dataset,” *ISMIR*, 2011.