

ユーザの生理状態に基づく楽曲の聴取選好の推定

馬 思宇[†] 牛尼 剛聡[‡]

[†]九州大学芸術工学府 〒815-0032 福岡県福岡市南区塩原 4-9-1

[‡]九州大学芸術工学研究院 〒815-0032 福岡県福岡市南区塩原 4-9-1

E-mail: [†] 2ds19002k@s.kyushu-u.ac.jp, [‡] ushiama@design.kyushu-u.ac.jp

あらまし 近年、膨大な楽曲の中からユーザに適した楽曲をする手法について活発に研究が行われている。しかし、従来の推薦手法の多くはユーザの嗜好が静的で不変であることを仮定している。しかし、ユーザが聞きたい音楽は、ユーザが置かれたコンテキストや内的な状態によって動的に変化する。つまり、たとえ同じ楽曲に対しても、ユーザが聴きたいときと聴きたくないときが存在する。そこで、我々は、ユーザは異なる生理状態で異なる聴取選好を持っているという仮説の下に、ユーザの生理状態と聴取選好の関係を調査した。本研究では、生理状態を測定できるセンシングデバイスを使用した実験実験と機械学習に基づいて、ユーザの生理状態と楽曲の物理的な特徴量との関連を抽出し、ユーザの生理状態に適した楽曲のプレイリストを構成する手法を提案する。

キーワード 楽曲推薦、生理情報、機械学習、聴取選好

1. はじめに

現在、社会の急速な発展に伴い、人々の生活のペースが速くなり、娯楽の重要性が増している。娯楽にはさまざまな種類が存在するが、その中でも音楽は、スマートフォン等の携帯型のデバイスを利用していつでもどこでも聴くことができるため、代表的な娯楽の一つになっている。そうした背景の下、音楽の利用環境に関する多くの研究が行われている[1]。

音楽に関する研究は様々な方向性に基づいて行われているが、音楽に関する研究として代表的な研究の一つに楽曲推薦に関する研究がある。これまでに、様々な楽曲推薦に関する手法が提案されている。例えば、ユーザベクトルの類似度に基づいて楽曲の推薦を行う手法[2][3]、楽曲レビュー或いはユーザの評価に基づいて類似度が高い楽曲を推薦する手法[4][5]、ユーザのプレイ履歴に基づいてユーザの嗜好を推定する手法[6]、歌詞の分析に基づいた楽曲の印象に基づく推薦[7]などがある。これらのほかにも多くの楽曲推薦システムが提案されている。

しかし、従来の研究では、ユーザの静的な状態だけを考慮している研究がほとんどであり、ユーザの嗜好状態の動的な変化を考えられていなかった。一般に、たとえ、どれほど好きな楽曲に対しても、ユーザは聴きたくないときがある。つまり、ユーザの気分によって聞きたいと思う楽曲は変わるものであるにもかかわらず、従来の研究では、このような問題を十分に考慮してこなかった。

この問題に対して、我々は「ユーザは異なる生理状

態で異なる聴取選好を持っている」という仮説立てた。本論文では、調査と予備実験によってこの仮説の妥当性を検証し、ユーザの生理状態の変化を考慮した楽曲推薦手法を提案する。

提案する手法では、ユーザが楽曲を聴きたいという気持ちを推定するために、ユーザの生理状態を取得し、機械学習によって推定する。本研究では、ユーザの生理状態の取得のためにはメガネ型センサを利用して、ユーザの生理状態が楽曲の聴取選好に及ぼす影響を検証し、リアルタイムの生理状態に基づいてプレイリストの自動構成と推薦を目標とする

2. 関連研究

2.1 プレイリストの推薦手法

棍ら[2]は、ユーザの視聴状況、歌詞と楽曲の情景という3つの特徴量を考慮し、特徴量空間モデルを構築し、楽曲の類似度を算出し、プレイリストを推薦する手法を提案している。

2.2 再生履歴に基づいた推薦手法

新美ら[6]は、再生履歴によりユーザの状況を推測し、楽曲を推薦する手法を提案している。この手法では、ユーザの過去の再生履歴と現在の再生履歴を比較することで、ユーザの気分の変化と各楽曲への評価を算出し、楽曲を推薦する。

2.3 音楽と生理に関する研究

大木ら[7]の研究では歌詞解析と心拍変動分析を用

いた楽曲による感情への影響の予備的な調査を行った。楽曲の歌詞に対して自然言語処理技術を用いて、感情分類を行い、それらの感情的な単語が視聴者の感情を喚起する結果を心拍数の変化に通じて示した。その原因はユーザの生理信号より、生理的な反応が出現していることを示した。

3. アプローチ

従来の音楽推薦手法ではユーザの楽曲再生履歴から類似する楽曲を推薦するのが一般的である。これらの手法では、ユーザの楽曲再生履歴からその時の気分に合わせて楽曲を推定できる。しかしそれらはあくまでユーザが音楽サービスを利用して、明示的にそのときに聞きたい楽曲を指定した後に限られる。音楽サービスを利用していない場合には、そのときにユーザが聞きたい楽曲を推定できないことが問題である。

そこで、本研究では、センサデバイスで計測されたユーザの動的な生理状態を元にユーザの聴取選好を推定し、そのときのユーザの状態に適したプレイリストを自動的に構成する手法を提案する。本研究で提案するシステムは、2つの部分から構成される。一つは、ユーザの生理状態とユーザが聞きたい曲の関係を機械学習モデルとして学習する部分であり、もう一つは、学習済みの機械学習モデルを利用して、ユーザの生理状態に基づく聴取選好の推定と楽曲の推薦リストの構成する部分である。

本研究では、機械学習のために、学習データとしてユーザの生理状態とユーザの楽曲に対する聴取選好の対応関係を利用する。ユーザの生理データはセンサデバイスで収集し、ユーザの各時点で曲の聴取選好データは、オーディオプレイヤーのログから取得する。そして、機械学習を使用して、ユーザの生理データと聴取選好データの対応関係を学習し、機械学習モデルを訓練する。学習が行われた後は、訓練された機械学習モデルにリアルタイムの生理データを入力すると、その時点のユーザの曲に対する聴取選好を推定し、推薦リストを構成する。

4. 予備実験

4.1 目的

本研究は、ユーザの生理状態が、ユーザが聞きたいと思う楽曲に影響を与えることを仮定としている。そこで、研究をすすめるにあたり、ユーザの生理状態がユーザの楽曲聴取の嗜好に与える影響に関する予備実験を行った。以下に、予備実験の内容を説明する。

4.2 予備実験の方法

予備実験では、センサデバイスによるユーザの生理状態の推定の誤差の影響を除去するために、被験者に対するアンケート調査により、ユーザの生理状態と思考する楽曲の影響の関連性の調査を行った。

・被験者：1人

・時間：2019年11月1日～2019年11月7日

まず、被験者に対して、被験者が普段から聴く頻度が高い上位20曲を選んでもらった。本研究の目的は、未知の楽曲からの推薦ではなく、ユーザが好きな曲を定説な状況で再生することにあるため、対象となるのは、全てユーザが好きな楽曲である。

次に、被験者に、毎日同じ時刻に、ユーザの生理状態と選んだそれぞれの楽曲に対する聴取選好を評価して記録してもらった。今回の予備実験では、生理状態を表す指標として「疲労度」と「テンション」という2つの指標を使用した。実験期間は1週間である。

評価指標の「テンション」は、-3から3までの値を取り、疲労度と聴取選好(楽曲に対する聴きたい程度)は0から4までの値を取る。テンションがマイナスの時は元気がない状態で、疲労が高い程、疲れているとする。聴取選好が0の曲は、ユーザが全く聴きたくないときで、4であるのはとても聴きたいときである。評価指標を表1に示す。

取得したデータに対して、相関分析を行った。分析した結果を表2に示す。その中、S1～S20は楽曲を表し、それぞれ疲労度とテンションの度合いとの相関係数を示している。

計算された相関係数表から見ると、少ないデータからにも聴取選好とパラメータの相関性が見える。

表1 評価指標

レベル	低い-----高い						
	-3	-2	-1	0	1	2	3
テンション							
疲労度	0	1	2	3	4		
態度	0	1	2	3	4		

表2 相関係数

	tiring 相関	tension 相関		tiring 相関	tension 相関
S1	-0.162	0.076	S11	0.103	0.311
S2	-0.269	0.539	S12	-0.092	0.508
S3	-0.194	0.522	S13	0.268	-0.147
S4	-0.282	0.639	S14	-0.385	0.612
S5	-0.102	0.486	S15	-0.385	0.364
S6	-0.372	0.231	S16	-0.373	0.543
S7	-0.220	0.169	S17	-0.190	0.511
S8	-0.125	0.265	S18	-0.319	0.610
S9	-0.236	0.570	S19	-0.177	0.581
S10	-0.176	0.448	S20	-0.053	0.435

4.3 予備実験の評価

相関係数の結果より、曲によって、プラスまたはマイナスの相関が強い楽曲が存在した。

疲労度との相関が一番高いのは S13 であり、ユーザが疲れたときには、聞きたい曲であることを表している。一方、S15 は疲労度と負の相関があり、疲労度が高いときには聞きたくない曲である。

テンションとの相関を考えると、S4 はユーザのテンションとの相関が高く、ユーザのテンションが高いときに聞きたくなる曲である。一方、S13 は、逆にテンションが低いときに聞きたくなる曲である。

実験結果を分布図にまとめたものを図 1~2 で示す。予備実験の結果より、ユーザの生理状態によって、ユーザの曲に対する嗜好に影響が与えられる可能性があることがわかった。

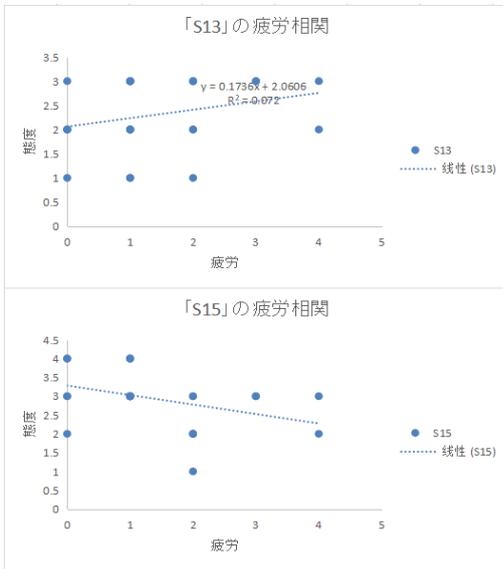


図 1 疲労相関散点図

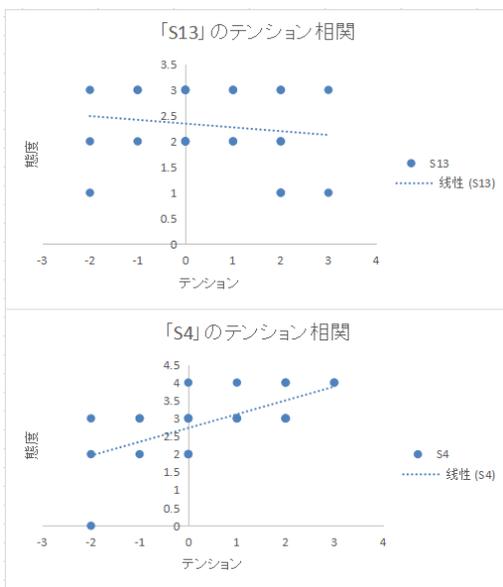


図 2 テンション相関散点図

5. 提案手法

本章では、ユーザの生理状態に基づいた楽曲のプレイリストの構成を行うための手法の概要を説明する。

本研究の提案手法は 4 つの手順がある。まずは機械学習を用いて、ユーザの生理状態と楽曲を聴く気分との関連性を学習する。そしてセンサーの利用より推定できるリアルタイムの生理状態を機械学習モデルへの入力データとして、ユーザのその時点の聴取選好を予測する。次は予測された聴取選好を元に入力とした楽曲のプレイリストをランキングし、上位の楽曲特を抽出する。最後に抽出された特徴量に基づいて類似する未知の楽曲を選別し、推薦用のプレイリストを生成する。

5.1 ユーザの生理情報の取得

5.1.1 生理状態の取得するセンサデバイス

ユーザの生理状態を測るために、本研究では JINS MEME というメガネ型ウェアラブルデバイスを使用する。本研究で使用するのは JINS 社の MEME ES (図 3) と呼ばれる製品であり、目電位センサーとモーションセンサーを搭載し、ユーザの頭と体の生理状態をリアルタイムで測定可能である (図 4)。

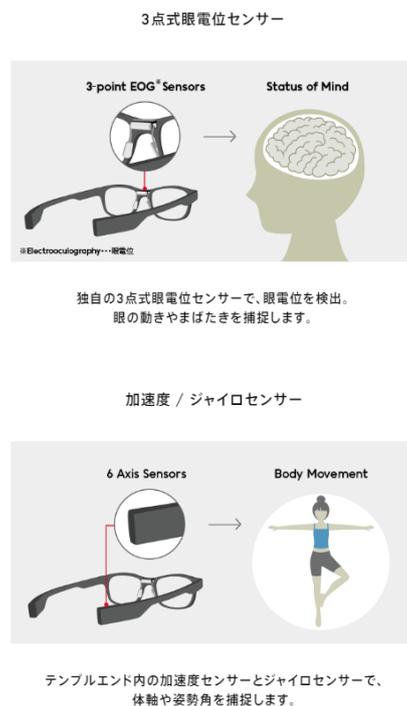


図 3 JINS MEME ES[17]



図 4 JINS MEME で生理情報を取得する

5.1.2 生理データを収集

JINS MEME から得られたデータを解析して、ユーザの生理状態を取得する。JINS MEME APP を利用すれば、リアルタイムで測られたユーザの生理状態を表示できる。

図 5 に実行例を示す。このアプリでは、ユーザの集中、活力、冷静、活動量、姿勢、安定性などな生理状態を測定可能であり、測定した生理状態より、利用者のリアルタイムの頭と体の状態を仮想年齢として示す。JINS MEME から提供された SDK (Software Development Kit) を利用すれば、それらの生理状態データを取得可能である。

- ・利用した APP : JINS MEME OFFICE
 - ・利用した API : advanced REST client
- 取得できるパラメータは表 3 のように示す。

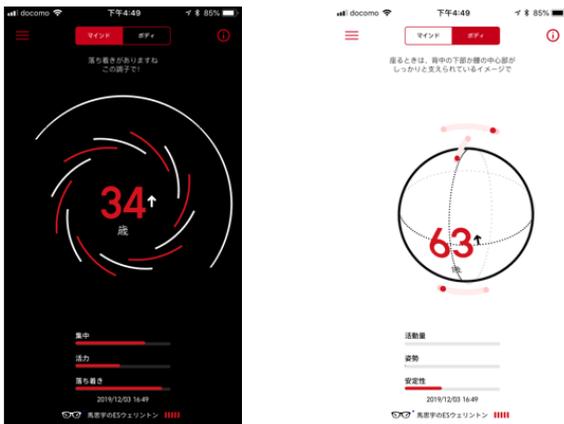


図 5 JINS MEME APP

表 3 取得可能なパラメータ

date	計測日時
zone	深い集中
focus	集中
clam	冷静
posture	姿勢
bki_sum	まばたきの間隔合計
bki_n	まばたきの数

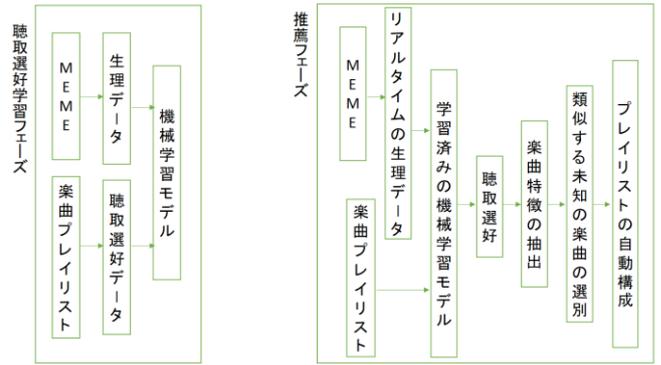


図 6 システムの概要

5.2 提案システム

本研究の提案システムは 2 つのフェーズに分けて説明する。提案システムの概念図は図 6 に示す。

5.2.1 聴取選好学習フェーズ

このフェーズは生理情報と聴取選好の関連性を構築するために、機械学習モデルを構築する。

データセットとしては 5.1 に紹介した通り、JINS MEME の利用によって、被験者の生理データと第 4 章に紹介した予備実験の手法と同じように収集した聴取選好データである(予備実験に収集したデータとは別)。それらのデータを訓練データとして機械学習モデルを訓練する。

今回、学習データと訓練データの割合は 4 : 1 で、5 段階の相互検証を実施した。

5.2.2 推薦フェーズ

このフェーズでは訓練された機械学習モデルを用いて、ユーザの気分にあうプレイリストを生成する。

入力データとしては JINS MEME の利用によって取得できるユーザのリアルタイムの生理データと機械学習モデルを訓練する際に提供されていた楽曲のプレイリストである。出力データとしては入力データとした楽曲プレイリスト中に全ての楽曲に対する聴取選好データ(予測値)になる。そしてそれらの聴取選好データに基づいて、プレイリスト中の楽曲をランキングする。ランキング上位の楽曲の特徴を抽出し、特徴ベクトルを作成する。最後、特徴ベクトルを元に楽曲のデータセットから類似度が高い上位の楽曲を選別し、新しいプレイリストを構成する。

5.3 楽曲の特徴抽出

今回は音楽とオーディオの分析用の Python パッケージとした Librosa の利用によって、楽曲の特徴を抽出する。抽出された楽曲の特徴は以下のように示す。

・クロマ周波数

クロマ周波数は、音楽オーディオの興味深く強力な表現であり、スペクトル全体が音楽オクターブの12の異なるセミトーン（または色度）を表す12の間隔に投影される。クロマ周波数は `librosa.feature.chroma_stft` で計算できる。

・メル周波数ケプストラム係数 (MFCC)

MFCCでは、人の聴覚特性に合わせて音素を変換し、スペクトルの包絡をフォルマント分析より多くの特徴量を用いて表現することで、フォルマント分析では捉えきれなかった音素の特徴を捉えることができる。主に音声認識の分野で使われることが多い。計 MFCC は `librosa.feature.mfcc` を利用して計算できる。

・スペクトル重心

スペクトル重心は音の「質量中心」がどこにあるかを示し、音に存在する周波数の加重平均として計算される。例えば、同じ長さのブルースの曲とメタルの曲比較すると、メタルの曲は最後の頻度が高くなる。したがって、ブルースの曲のスペクトル重心はそのスペクトルの中央近くに位置され、そしてメタルの曲のスペクトル重心は、中心からずれる。スペクトル重は `librosa.feature.spectral_centroid` をりようして計算できる。

・ロールオフ頻度

ロールオフ頻度は、信号形状の尺度である。ローフオフ頻度は `librosa.feature.spectral_rolloff` により計算できる。

・ゼロクロッシングレート

ゼロクロッシングレートは音声認識や音楽情報の取得に広く使用されている。金属や岩のような衝撃の強い音の場合、通常はより高い値になる。ゼロクロッシングレートは `librosa.feature.zero_crossing_rate` を利用して計算できる。

6. 評価実験

本実験の提案手法を検証するために、回帰分析に基づいた評価実験を行った。

6.1 実験内容

本実験の流れとしては、4章で示した予備実験のように、被験者の聴取選好データを収集した。その同時に、JINS MEMEの利用より、生理データも収集した。

- ・被験者：1人
- ・器材：JINS MEMES
- ・時間：2020年5月24日～2020年6月2日
- ・生理データ：30分ごとに一回平均の取得
- ・聴取選好データ：30分ごとに一回記録
- ・データ数：34個

表4 データサンプル

date	zone	focus	calm	posture	bki_sum	bki_n
22:00	66.01	86.66	46.02	75.84	14.50	2.89
21:30	62.57	84.94	40.85	69.67	14.85	3.35
21:00	59.23	84.88	34.00	74.58	14.47	2.45
20:30	60.36	79.35	42.03	71.92	14.41	3.14
20:00	65.56	77.72	53.98	59.28	14.28	5.11
19:30	61.78	90.14	33.90	64.84	15.03	2.39
19:00	44.18	54.13	34.79	76.55	14.50	1.08
18:30	26.77	9.99	43.78	81.93	14.89	0.70
18:00	50.70	36.03	65.68	71.26	14.78	1.53

今回利用した生理データは集中度、冷静度、姿勢を表す `zone`、`focus`、`calm`、`posture` という四つのパラメータのみである。まばたきの規則性が低いため今回は考慮しなかった。データのサンプルを表4に示す。

6.2 回帰モデル

今回の回帰評価に使用したのは 'BayesianRidge', 'LinearRegression', 'ElasticNet', 'SVR', 'GBR' という五つの回帰モデルである。

6.3 分析評価

データ数が少ないため、本実験では訓練データと検証データ（正解データ）が二つの割合に分けて評価分析を行った。

・ `test_rate=0.5` の場合（訓練データ：検証データ = 1：1）

例とする曲 S20 の聴取選好予測曲線と散点図は図7、図8に示す。

評価分析は表5に示す。

・ `test_rate=0.2` の場合（訓練データ：検証データ = 4:1）

例とする曲 S20 の聴取選好予測曲線と散点図は図9、図10に示す。

評価分析は表6に示す。

分析結果からみると、`test_rate=0.2` の場合に BayesianRidge と ElasticNet の予測結果がより良く、`test_rate=0.5` の場合に BayesianRidge の予測が一番良かった。

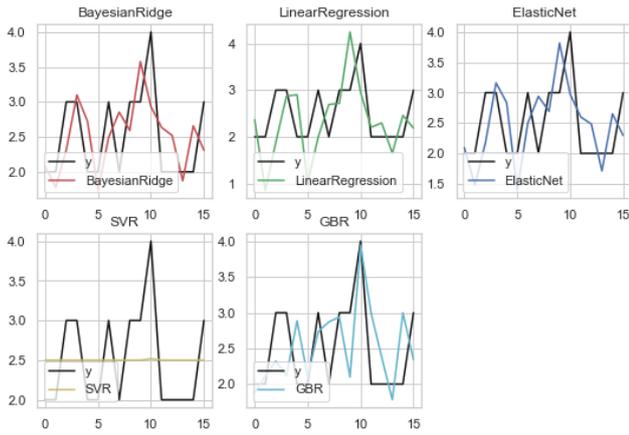


図 7 曲 S20 の聴取選好予測曲線 (test_rate=0.5)

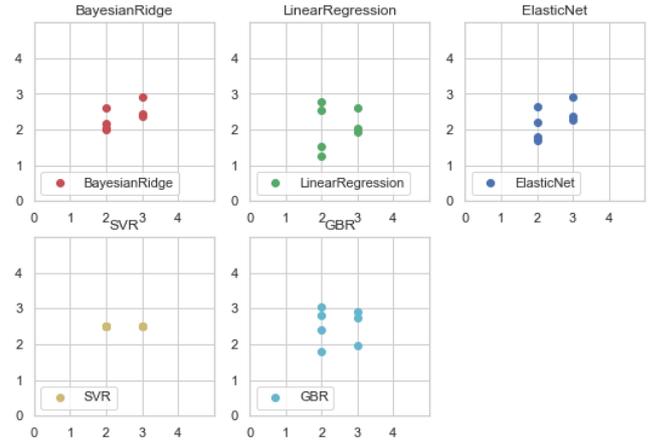


図 10 曲 S20 の予測値と正解値の散点 (test_rate=0.2)

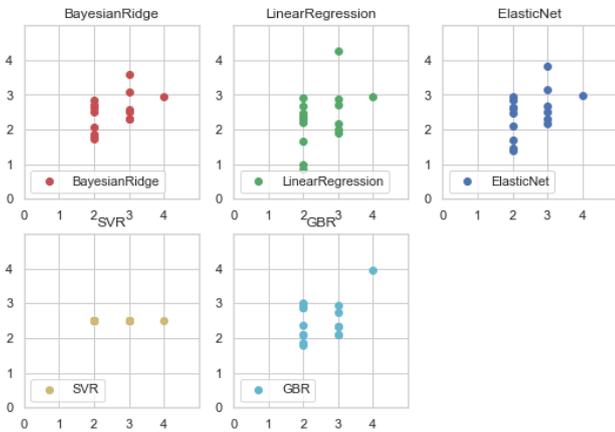


図 8 曲 S20 の予測値と正解値の散点 (test_rate=0.5)

表 5 回帰分析評価 (test_rate=0.5)

test_rate=0.5	ev	mae	mse	r2
BayesianRidge	0.05	0.52	0.41	-0.03
LinearRegression	-0.28	0.6	0.56	-0.38
ElasticNet	-0.03	0.54	0.44	-0.09
SVR	0	0.55	0.43	-0.08
GBR	-0.33	0.56	0.57	-0.42

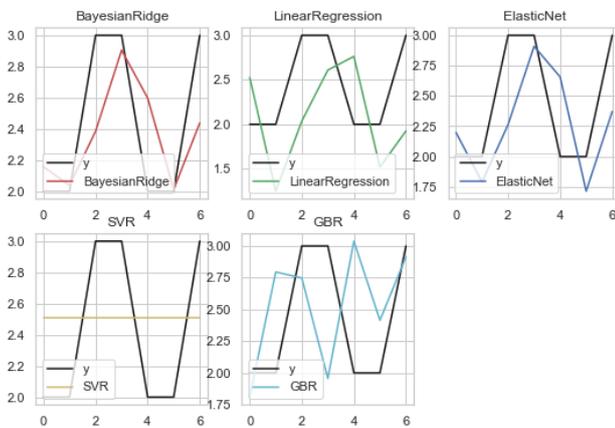


図 9 曲 S20 の聴取選好予測曲線 (test_rate=0.2)

表 6 回帰分析評価 (test_rate=0.2)

test_rate=0.2	ev	mae	mse	r2
BayesianRidge	0.012	0.53	0.4	-0.1
LinearRegression	-0.14	0.53	0.43	-0.28
ElasticNet	0.09	0.52	0.39	-0.07
SVR	0	0.6	0.49	-0.37
GBR	-0.68	0.67	0.68	-1.04

6.4 実験評価

本実験では五つの回帰モデルを用いて評価分析を行ったが、今回の実験には以下に示す不十分な点が存在する。

- ・被験者が少ないため、一般性が足りない。
- ・データの数が少ないため、学習効率が高くない。
- ・アンケート調査で行われたため、記録した聴取選好は主観的で、客観性が足りない。

今回の実験では、ある程度ユーザの聴取選好を推定することができる可能性を示すことができたが、精度が高くない。本研究の提案手法をもっと正確に評価するために、今後、さらに実験を行う必要がある。

7. おわりに

本論文では、ユーザの生理状態が楽曲に対する聴取選好に与える影響を討論し、予備実験の設置を通じて仮説を検証した。そして、ユーザの生理状態に基づいて、機械学習を利用し、プレイリストを構成する手法を提案した。

今回の評価実験ではユーザ主観的な聴取選好のデータを使用したため、データの誤差が大きいと考えられる。より正確に聴取選好のデータを抽出するために、今後はユーザの再生履歴から楽曲への聴取選好を推定する手法を考えている。

ユーザの聴取選好は人より異なるものである。提案したシステムではデータを収集した対象のユーザしか適用できないという欠点がある。今後は複数の被験者

に適応する実験の実施を考慮し、対象ユーザを一般的にする推薦手法の提案及び推薦システムの開発を考えている。

[17] ©JINS MEME,” JINS MEM DEVELOPERS”,
<https://jins-meme.com/ja/developers/>

謝辞

本研究は JSPS 科研費 19H04219 の助成を受けたものです。

参 考 文 献

- [1] 吉井和佳, 後藤真孝, “音楽情報処理技術の最前線: 7. 音楽推薦システム”, *情報処理*, Vol.50, No.8, pp.751-755, 2009.
- [2] 梶克彦, 平田圭二, 長尾確, “状況と嗜好に関するアノテーションに基づくオンライン楽曲推薦システム”, *National Convention*, pp.21-22, 2005.
- [3] 永田裕二, 帆足啓一郎, 菅谷史昭, 甲藤 二郎, “複数ユーザの音楽嗜好特徴空間を用いた楽曲推薦システム”, *National Convention*, pp.173-174, 2006.
- [4] 吉井和佳, 後藤真孝, 駒谷和範, 尾形哲也, 奥乃博, “ユーザの評価と音響的特徴との確率的統合に基づくハイブリッド型楽曲推薦システム”, *MUS*, pp.45-52, 2006.
- [5] 水野智公, 亀谷由隆, “単語のベクトル表現に基づき楽曲要素の足し引きを行う音楽推薦”, *National Convention*, pp.345-346, 2019.
- [6] 新美怜志, 濱川礼, “再生履歴によるユーザの状況推測に基づく音楽推薦システム”, *MUS*, pp.1-4, 2009.
- [7] 大木麻里衣, 丸野由希, 久保孝富, “歌詞解析と心拍変動分析を用いた楽曲による感情への影響の予備的調査”, *情報処理学会関西支部 支部大会講演論文集*, 2018.
- [8] 奥健太, “セレンディピティ指向情報推薦の研究動向”, *知能と情報* Vol.25, No.1, pp.2-10, 2013.
- [9] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, “Audio Chord Recognition with Recurrent Neural Networks”, *ISMIR*, pp.335-340, 2013.
- [10] 古屋瑞生, 黄宏軒, 川越 恭二, “歌詞情報に基づく聴取目的に応じた楽曲推薦システムの提案”, *National Convention*, pp.625-626, 2014.
- [11] 吉井和佳, “音楽を軸に広がる情報科学: 5. 音楽と機械学習”, *情報処理*, Vol.57, No.6, pp.519-522, 2016.
- [12] 亀岡弘和, “音楽を軸に広がる情報科学: 1. 音楽と信号処理”, *情報処理*, Vol.57, No.6, pp.506-509, 2016.
- [13] D. Mekara, and V. Gupta, “SCDV, Sparse composite Document Vectors using soft clustering over distributional representations”, *Proc. of EMNLP*, 2017.
- [14] 宮原桃子, “眼の動きだけで、音楽と映像を奏でる!? ALS 患者の未来を変える世界初のミュージックフィルムに託す夢”, <https://soar-world.com/2018/05/08/jinsmeme/>, 2018.
- [15] @komde, “JINS MEME API で生体データ可視化にチャレンジ!”, <https://qiita.com/komde/items/79c4ff4b29685cbc6c21>, 2018.
- [16] JINS MEME,” アプリの開発を始める”, <https://jins-meme.github.io/sdkdoc/>.