

曖昧な印象を内包する空間画像をクエリとする音楽レコメンドの提案

Proposal of music recommendation using spatial images containing ambiguous impressions as a query

○石橋 京^{*1}, 山田 悟史^{*2}

Kyo ISHIBASHI^{*1}, Satoshi YAMADA^{*2}

*1 立命館大学 理工学部 建築都市デザイン学科

Undergraduate, Department. of Architecture and Urban Design, Ritsumeikan University.

*2 立命館大学 理工学部 建築都市デザイン学科 准教授・博士 (工学)

Associate Professor. of Architecture and Urban Design, Ritsumeikan University. Dr.Eng

キーワード：レコメンドシステム；音楽；意外性

Keywords: recommendation system; music; unexpectedness.

1. 背景

近年、情報技術の発達やインターネット普及によって人々は様々なサービスを利用できるようになっている。その中の一つに、Netflix¹⁾やAmazon²⁾がある。これらのサービスの多くには、ユーザ自身による検索とシステムによるレコメンドがある。検索では言語が多く用いられ、ユーザ自身による希望の言語化が必要である。一方でユーザ自身でも希望の言語化が困難な場合もある。希望が曖昧かつ言語化に適さない場合である。「可愛い、明るい、爽やかな」などの印象が代表的である。このような場合の対応として、レコメンドが併用されている。ユーザの履歴をもとにするシステムが多く、代表的なアルゴリズムには協調フィルタリングがある。多くのサービスに実装されている有効なアルゴリズムである。一方で、履歴に対する過度なフィッティングという課題もある。この課題では、レコメンド内容に偏りが生じ、ユーザが言語化困難な曖昧・潜在的な希望に対応する目的が十分には果たせない可能性が指摘されている。このような課題解決は、妥当性と意外性を併せ持つコンテンツとの出会いを促すという意味で生活満足度の向上に資すると考えられる。筆者が知る限りこのようなシステムを持つサービスは見当たらず、研究も少ない。

以上から本研究は、ユーザに曖昧な希望の言語化を要求しないシステムを提案する。具体的には曖昧な印象を内包する画像をクエリとするシステムである。題材は音楽視聴とする。

2. 関連研究

意外性を十分に考慮したレコメンドシステムの関連研究として、梶谷氏によるユーザと他人の感性を取り入れたファッションレコメンドモデルなどがあげられる。³⁾⁴⁾

この研究では、ユーザ個人の感性モデルのみを取り入れた従来のファッションレコメンドでは、ユーザが好ん

で着ているファッションと同系統のものしか推薦されないことや、自身が選んだアイテムからレコメンドアイテムが選ばれるために、自身の感性に自信がないユーザにとって不便であることを問題点として挙げた。それらの解決策として他人の感性モデルも取り入れることで、自身の感性のみを取り入れたレコメンドモデルよりも様々なアイテムがレコメンドされるようになり意外性が増し、また自身のセンスに自信がない人でも扱いやすくなることを目標とした。

提案モデルで取り入れられたアルゴリズムでは、ユーザの普段着に似ている画像と、WEAR⁶⁾というファッションサイトから収集できる他人のアカウントごとの画像をそれぞれ教師データとしてk-means法によるクラスタリングをおこない、ユーザのクラスタと他人のクラスタ2つとのユークリッド距離の総和が小さいものをレコメンドする、自身と他人の2つの感性を取り入れたモデルを提案した。個人と他人の両クラスタとのユークリッド距離の総和を指標とすることで、自身のクラスタと他人のクラスタの間に位置するものがレコメンドされやすくなり、自身の趣味趣向とは離れ過ぎずに他人の感性モデルも反映させられることが可能となった。

その結果、提案モデルであるユーザと他人の感性両方を取り入れたモデルによる意外性の向上が見られ、提案モデルの有用性が示唆された。

以上をふまえて本研究は、曖昧な印象を内包する画像をクエリとする音楽レコメンドの提案と検証を目的とする。類似の提案と検証をした研究を見当たらず、この点が本研究の新規性に該当する。なお提案システムを既存システムに合わせて端的に呼称すれば画像検索であるが、抽象度が高い検索であることから本研究では、レコメンドと呼称する。

3. 研究概要

本研究では曖昧な印象を内包した空間画像をクエリとしてユーザに音楽をRecommendするシステムの提案と検証を行う。

3.1. 提案システム

提案システムでは、ユーザの周りの空間からその空間に適した音楽をRecommendする。具体的には、ユーザが選んだ空間画像をシステムに渡すことで、ジャケット画像とクエリ画像との類似度を計測し、最も類似度が高いジャケット画像とその音楽をRecommendするものである。

クエリ画像とジャケット画像の類似度を測る指標として本研究では①画像のRGBの類似度②物体検出による画像に写る物体群の類似度の2つを用いた。

3.1.1. RGBの類似度の計測方法

RGB各色の0~255階調のピクセルの数をヒストグラムとして表し、それをベクトルに直したもので相関係数を算出する。例を図1に示す。

3.1.2. 物体検出の類似度の計測方法

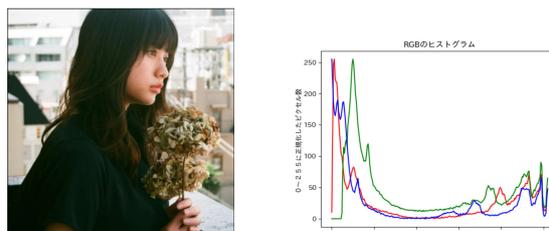
物体検出モデルとしてfaster R-CNNを用いた。検出した物体のconfidenceが0.7以上であるものをカウントし、それらのカウント数から相関係数を算出する。なお、空間画像もしくはジャケット画像の物体検出数がすべての物体で0である場合は相関係数が算出できないため、代わりに相関係数を0として扱う。例を図2に示す。

3.2. 検証実験

提案システムが空間に適した音楽をRecommendできているかを確認するために被験者実験を行った。実験では、被験者に明るい空間または暗い空間画像を選んでもらい、そのクエリ画像から選ばれたジャケット画像とその音楽をRecommendする。空間画像を明暗で分けた理由は、空間の特徴によってRecommendの良否が変わるのかを分析するためである。その後音楽を実際に聴いてもらい、用意したアンケートに答えてもらった。実験対象は20~27歳の男性24名、女性6名の計30名とした。

3.2.1. 音楽

本実験では、音楽再生アプリ「Spotify」⁷⁾からとることのできる音楽データを用いた。Spotifyにはアーティスト情報やジャケット画像のURLだけでなく、曲のテンポやポジティブかネガティブかを判別するなど数多くのメタデータが存在し、それらを用いて実験と分析を行った。システムに用いる音楽はJ-POPを中心とした1990年代~2020年代の楽曲600曲とした。また、ベストアルバムな



曲名: ドライフラワー アーティスト: 倭斐

Spotify: <https://i.scdn.co/image/ab67616d0000b273a89f91a19c16e94ba441132> (2023/10/27)

図1 RGBのヒストグラム計測例



曲名: Do not touch アーティスト: MISAMO

Spotify: <https://i.scdn.co/image/ab67616d0000b273a50946cb5f2ade7894904e73> (2023/10/27)

background	person	bicycle	car	motorcycle	airplane	bus	train	truck	...	toaster	sink	refrigerator	book
0	3	0	0	0	0	0	0	0	...	0	0	0	0

図2 物体検出の計測例

どはジャケット画像が曲の雰囲気や反映させづらいという観点から、今回扱う楽曲はシングル曲とそのジャケット画像のみとした。

3.2.2. 質問

Recommendされたジャケット画像とその曲について、
・Q1.ジャケット画像がその曲の雰囲気や反映しているか
・Q2.ジャケット画像と空間画像が似ているか
・Q3.空間画像にその曲が適しているか
の3つの質問に5件法で回答してもらった。

3.2.3. 提案モデル

提案システムが有用であるかを確認するために本実験では以下の4つのモデルでRecommendを行い、それぞれのモデルのアンケート評価を比較した。

- ・モデル1.ランダムに楽曲をRecommendするダミーモデル
- ・モデル2.RGB比較を用いたモデル
- ・モデル3.物体検出のみを用いたモデル
- ・モデル4.RGB比較+物体検出を用いたモデル

各モデル1名あたり暗い空間2回、明るい空間2回の合計1人あたり16回のRecommendを行った。なお、被験者にはRecommendのアルゴリズムやどのモデルでRecommendを行っているかを知らせていない。あるクエリ画像一枚に対するモデル2~4のRecommendの例を図3に示す。なお、実際にユーザに表示されるのはジャケット画像、曲名、アーティスト名のみである。

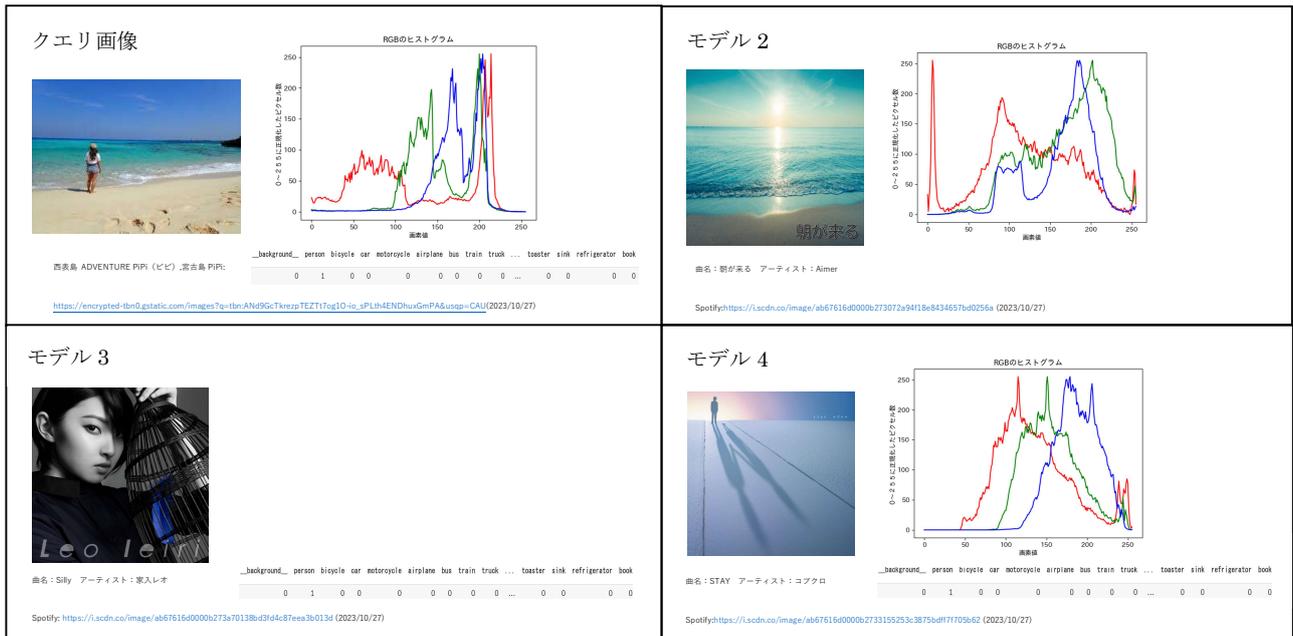


図3 クエリ画像に対するモデル2~4のレコメンド例

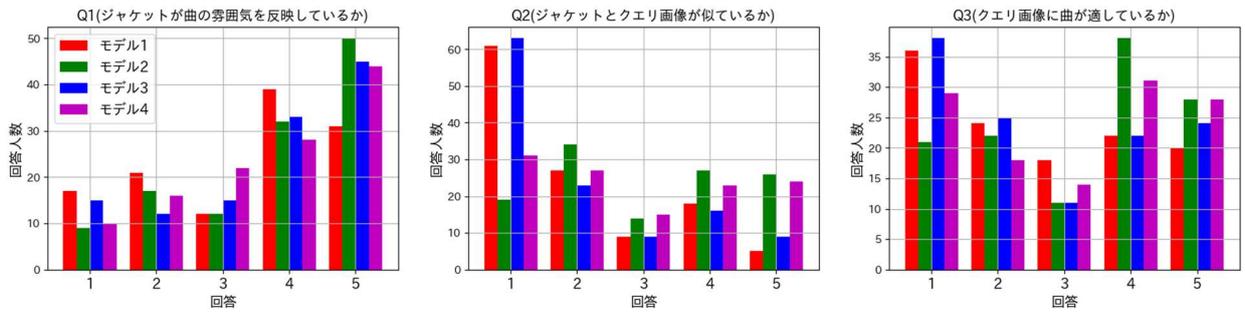


図4 各アンケート項目の集計

表1 各モデルのアンケートの平均

	Q2	Q3	I	II	III
モデル1	1.94	2.66	0.18	0.35	0.75
モデル2	2.99	3.13	0.43	0.52	0.73
モデル3	2.29	2.96	0.29	0.42	0.75
モデル4	2.81	3.30	0.38	0.51	0.73

表2 空間の明るさごとのモデル2.4の評価平均

		Q2	Q3	I	II
暗い空間	モデル2	3.06	2.92	0.44	0.46
	モデル4	2.90	2.86	0.38	0.29
明るい空間	モデル2	2.92	3.38	0.41	0.59
	モデル4	2.75	3.59	0.38	0.66

- ※ I. Q2 が 4 以上の割合
- II. Q3 が 4 以上の割合
- III. Q1 が 4 以上かつ Q2 が 4 以上での Q3 が 4 以上の割合

- ※ I. Q3 が 4 以上の割合
- II. Q1 が 4 以上かつ Q2 が 4 以上での Q3 が 4 以上の割合

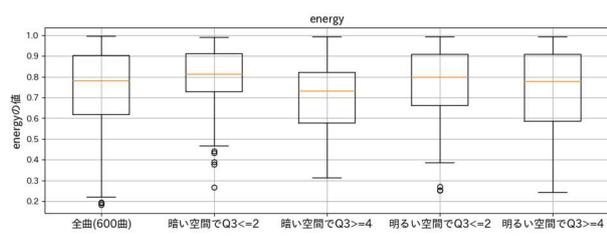


図5 空間の明るさごとのメタデータの比較(energy)

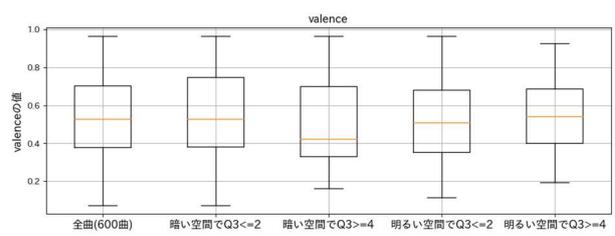


図6 空間の明るさごとのメタデータの比較(valence)

4. 結果と考察

各モデルのアンケート回答の実験の結果を図4に示す。Q2、Q3では物体を検出できていないレコメンドを含むモデル3と、モデル1の評価が低く、モデル2と4の評価が高かった。

以降の結果は、モデル1と、モデル3で物体を検出できずにレコメンドしたものを合わせてモデル1と呼称し、モデル3は物体を検出したレコメンドのみを扱う。さらにモデル4で物体を検出できずに色の比較のみでレコメンドしたものをモデル2に含み、モデル4は物体を検出してさらに色の比較も行ったレコメンドのみを扱う。

4.1.各モデルのアンケートの結果

Q2.Q3の平均値に加えて、各モデルの次の3つの指標の評価値を表1に示す。

I. Q2の回答が4以上の割合

II. Q3の回答が4以上の割合

III. Q1.Q2の回答がともに4以上のうちのQ3の回答が4以上の割合

モデル2-4すべてでモデル1よりも似たジャケット画像と空間に適した音楽を推薦できていることがわかる。モデル3は、Q2がモデル2とモデル4に比べて低い結果となった。これは、クエリとジャケットで同じ画像が写っていても、画像が似ていないというユーザの評価が多いからだと考える。図7に例を示す。2つとも人を検出しているが、画像が似ているとは言えない。また、IIIの結果がどのモデルでも0.75と高いことから、IIの結果を高めることが空間に適した音楽を推薦することにつながる事がわかる。

4.2.空間の明るさごとのRGB比較の評価

空間の明るさによって受けるRGB比較によるレコメンドの評価の違いを表2に示す。

明るい空間のほうが画像の類似度の評価は少し低いものの、Q3の結果は高いことがわかる。

4.3.メタデータの比較

空間の明るさによってどのような曲のレコメンドが有効であったかを分析するためにSpotifyで取得できるメタデータを分析した。今回分析に用いるメタデータは、

- ・energy:エネルギー感(0~1)
- ・valence:曲のポジティブ度(0に近ければネガティブ、1に近ければポジティブ)

の2つとした。これら2つを選んだ理由は、曲の明暗をはっきりと分ける指標と考えたからである。クエリ画像の明暗、Q3の回答別に比較した結果を図5に示す。Energyのグラフを見ると、暗い空間で評価が低かったものはenergyが高く、評価が高かったものはenergyが低いことがわか



西表島 ADVENTURE PIPPI (ピピ) 演古島 PIPPI: クエリ画像

<https://mijako-pipi.com/pipicms/wp-content/uploads/2021/09/B95EBBC>

C67A-41D5-8879-BFA7AFBD447-1289-960.jpg (2023/10/27)



曲名: 標、アーティスト: 佐藤紗花 レコメンド画像

Spotify:

<https://a.scdn.co/image/ab67616d0000b771454463b61bb9446c6c7d> (2023/10/27)

図7 モデル3のレコメンド失敗例

る。明るい空間で評価が高いものはenergyが低いものから高いものまでさまざまである。Valenceのグラフを見ると、暗い空間で評価が高かったものはネガティブな曲が多いことがわかる。これらから、明るい空間では元気な曲から落ち着いた曲まで幅広くユーザが聞きたがっていて、表3のように明るい空間でのQ3の結果が高かったと考えられる。

5. まとめ

本研究では妥当性と意外性を併せ持つコンテンツとの出会いを促すことを目的とした、空間画像から音楽を推薦するシステムの開発を試みた。空間に似たジャケット画像を推薦できれば空間に適した音楽に出会うことを促せられることが分かった。今回用いたRGBの類似度によるレコメンドはユーザの選んだ空間に適した音楽を、ランダムに推薦するよりも高い割合で推薦できた。一方で物体検出を用いたレコメンドでは、画像の類似度が低く空間に適した曲が推薦されないなど課題が残った。今後は物体検出のアルゴリズムや類似度算出の方法を改善することが必要である。本研究を進めることで、妥当性と意外性を併せ持つ音楽のレコメンドだけでなく、建築物や飲食店、見ている景観などに合わせた音楽を自動で再生することも可能となり、今後の生活の質向上を示唆している。

[参考文献]

- 1) Netflix : <https://www.netflix.com/jp/>(参照 2023-09-27)
- 2) Amazon : <https://www.amazon.co.jp/>(参照 2023-09-27)
- 3) 梶谷 知嗣 : ユーザと他人の感性モデルを利用したファッションレコメンド、大学院研究年報 理工学研究科編、p.46(2016)
- 4) 土田滉也 : セレンディピティを考慮した推薦システムの提案、第79回全国大会講演論文集、p473-474(2017)
- 5) 瀧山幸大、藤田邦彦 : 音楽を聴いて色を連想する共感覚を利用した楽曲の推薦、第81回全国大会講演論文集、p347-348(2019)
- 6) WEAR : <https://wear.jp/>(参照 2023-09-27)
- 7) Spotify : <https://www.spotify.com/jp/free/>(参照 2023-09-27)