The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

場の雰囲気を考慮したBGM推薦システム構築の試み

BGM Recommendation System Considering the Atmosphere

佐藤 季久恵 *1 Kikue Sato　坂井 栞 *1 Shiori Sakai　高屋 英知 *2 Eichi Takaya　山内 和樹 *3 Kazuki Yamauchi　大矢 隼士 *3 Hayato Oya　栗原 聡 *2 Satoshi Kurihara

*1電気通信大学 大学院情報理工学研究科
Graduate School of Informatics and Engineering, The University of Electro-Communications

*2慶應義塾大学 大学院理工学研究科
Graduate School of Science and Technology, Keio University

*3株式会社レコチョク
RecoChoku Co.,Ltd.

Japanese sometimes leave the decision of things to the atmosphere, it is very important to design the atmosphere of the place in various situations. BGM has an emotion inducing effect and an image inducing effect, which makes it possible to change the atmosphere of the space. In this paper, we proposed a BGM recommendation system reflecting the atmosphere in the shop by combining images and environmental sounds and constructing a data set using shop labels. As a result of recommendation by our system, higher ratings than baseline were obtained for three index.

1. はじめに

場の雰囲気は，人と人との間に流れる「空気」や視覚情報，聴覚情報などが複雑に絡み合うことにより形成される．これを人為的にデザインすることを考えたとき，特に視覚情報と聴覚情報をコントロールすることが重要である．家具や調度品，壁紙など，BGM には感情誘導効果やイメージ誘導効果があるとされており [1]，場の雰囲気を小さい労力で変容させることができる．現にレストランなどでは，それらの効果を利用するべく，楽曲配信サービスを導入しているところが多い．しかし，時間帯による店内の見え方や客層の変化により，店舗側はその都度雰囲気に合った BGM を提供することが求められる．時々刻々と変化する店内の状況に合わせて BGM 配信サービスが対応したり，人手による選曲を行うことは困難であることから，本研究では，店舗の雰囲気に適した BGM を推薦するシステムの構築を行う．本稿では，あらかじめ撮影された店舗内動画と，異なる環境音を組み合わせて印象評価を行い，類似度に基づいた楽曲推薦を試みる．

2. 関連研究

画像や楽曲といったコンテンツに対しユーザがタグ付けを行うことは，ソーシャルタギングと呼ばれ，これを利用した推薦や検索システムに関する研究は数多く存在する．梶ら [2] は，歌詞とアノテーションを利用し，視聴時のユーザの状況に合わせたプレイリストを作成するために楽曲とユーザを特徴量空間へマップする手法を採用している．特徴量には歌詞，楽曲情景，視聴状況を用いており，それらの特徴量空間にユーザをマップすることで，楽曲間，ユーザと楽曲間，ユーザ間の類似度計算を可能にしている．楽曲情景のラベルは登場人物（一人，私），いつ（朝，過去，春），状況（恋愛中，反社会），心理状況（悲しい，怒り）の 4 項目を用いている．歌詞と楽曲情景については，それまで視聴した好きな曲の特徴量平均をそれぞれの特徴量空間にそのユーザの嗜好としてマップすること

連絡先: 佐藤季久恵，電気通信大学大学院情報理工学研究科情報学専攻，東京都調布市調布ヶ丘 1-5-1，ksato@ni.is.uec.ac.jp

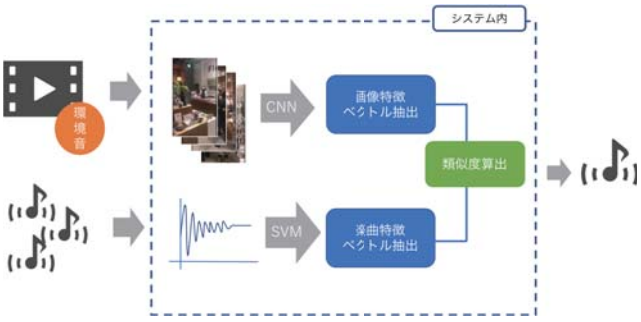

図 1: システムの概要図

で．推薦を行っている．Kaminskas ら [3] の研究ではユーザが関心のある場所（place of interest, POI）に即した楽曲推薦のシステムを構築している．楽曲と POI に双方同様の感情語を用いたタグをつけ，それらをベクトルとして扱い，類似度から適した楽曲を推薦している．タグには 9 項目の感情タグと 13 項目の物理的タグ（色や気温など）を用いている．

3. 本研究のアプローチ

本研究では，店舗での利用を想定し，場の雰囲気に合った BGM の推薦システムを提案する．

3.1 システムの概要

本研究における場の雰囲気を反映した BGM 推薦の流れを図 1 に示す．

まず，店舗内動画像を提案システムに入力し，Convolutional Neural Network（CNN）によってラベルを予測する．そして，推薦の候補となる楽曲群から特徴量としてメル周波数ケプストラム係数（MFCC）を求め，それを入力とした Support Vector Machine（SVM）を用いてラベルを予測する．それぞれの楽曲と店舗内動画において予測されたラベルの類似度を算出し，類似度が高いものを店舗に適した楽曲として推薦する．

3.2 環境音の利用

先行研究の多くは，場所の画像に対してラベル付けがなされている．画像データの取得やラベル付けは容易である一方，推薦システムにおいては視覚情報しか考慮されない点で不十分である．そこで本研究では，動画を印象評価に用い，その中でも環境音に着目する．

環境音は，我々が普段生活する中で無意識に耳にし，そこがどのような場所であるのかを判断するための材料にもなりうる．また，環境音が存在することにより，場所の見え方も変化する傾向があるとされている [4]．本研究では BGM を独立した音として扱うのではなく，環境音を含めて場をデザインすることを考慮し，環境音を含めた印象評価を行う．

3.3 店舗ラベルの導入

感情語や印象語を用いて画像や楽曲にラベル付けを行い，それに基づいた楽曲推薦を行うことは，先行研究でも試みられている．ラベルを用いることで，画像や楽曲がユーザーへ与える心理的影響を考慮することができ，個人の嗜好や状況に即した楽曲などを推薦することが可能である．しかし，個人の経験や感情に依存する感情語や印象語ラベルを使用することは不特定多数への適応を目的とした BGM 推薦には不向きである．

そこで本研究では，画像を見て想起される「情景」として，スターバックスや東急ハンズといった具体的な店舗名を用いてラベル付けを行う．チェーン店は全体を通してコンセプトを持っており，店舗内装などを統一しているところが多い．そのため，店を利用した人の間では同一のイメージを共有することが可能である．また，これまでは複数の感情語や印象語を用いることで店舗内を表現していたが，具体的な店名を用いることで一つのラベルで表現することが可能になる．

4. データセット

4.1 ラベルの選定

先述した店舗ラベルの候補を選定するにあたり，聞き取り調査を行った．

回答者は学生 4 人，社会人 2 名で，店舗イメージが確立している店舗名を列記してもらった．集計後，店の種類に偏りが出ないよう，USEN のコンシェルジュサービス [5] を参考に記載されている項目で補った．ラベル内容は表 1 に示す．これらのラベルを，後述する店舗内動画データセットの作成と楽曲評価に用いた．

4.2 店舗内動画データセットの作成

店舗内の雰囲気を評価するための店舗内動画を収集した．撮影は目線の高さで店舗内を 180 度撮影とし．時間は 10 秒から 15 秒ほどで行った．画像サイズは 1920 × 1080，フレーム数は 30fps とした．撮影イメージを図 2 に示す．ライトやアプリなどで明度や彩度の調整はしないものとした．

収集された店舗内動画から環境音を切り離し，異なる環境音を付け加えることで仮想店舗内動画を作成した．環境音はカフェ店内，子供が多いイベント会場内，ショッピングモール内，オフィス内，街中の 5 種類を使用した．

4.2.1 店舗内動画評価

作成した 740 本の仮想店舗内動画に対し，3 名の作業者に印象評価を行ってもらった．具体的には，作成した仮想店舗内動画を視聴してもらい，動画がどのような状況に当てはまるか，表 1 のラベルから選択してもらった．

表 1: ラベル一覧

情景	
スターバックス	ダイソー
ルノアール	東急ハンズ
コメダ珈琲	無印良品
バー	ニトリ
居酒屋	大塚家具
割烹・料亭	紀伊國屋書店
ラーメン屋	TSUTAYA
沖縄料理店	高島屋
イタリア料理店	PARCO
西友	イオンモール
成城石井	シェラトンホテル
カルディ	オフィス
アパレル（高級店）	企業ロビー（大企業）
アパレル（フォーマル）	企業ロビー（中小企業）
アパレル（カジュアル）	国際空港
ドラッグストア	地方空港
francfranc	
日差し	
あり	なし
時間帯	
朝（-10 時台）	昼（11 時台-14 時台）
夕方（15 時台-17 時台）	夜（18 時台-）
都市度合い	
都会	郊外
田舎	

4.3 楽曲評価

仮想店舗内動画と同様に，多数の楽曲に対して，それらがどのような状況下で BGM として流れているかを基準に，3 名の作業者に評価を行ってもらった．楽曲は「 J-POP」,「アニメ」,「キッズ・ファミリー」,「歌謡曲・演歌」,「邦楽ヒップホップ・R & B・レゲエ」,「邦楽ロック」,「洋楽ヒップホップ・R & B・レゲエ」,「洋楽ポップス」,「洋楽ロック」,「洋楽総合」の計 10 個のジャンルから 100 曲ずつを用意し，ラベルは表 1 から選択してもらった．

5. 予測器の構築と推薦方法

前節で作成したデータセットを用い，店舗内動画および楽曲に対する予測ラベルを出力するための予測器をそれぞれ構築した．

店舗内動画から予測ラベルを出力する予測器には，CNN を用いた．ネットワーク構造は AlexNet[6] と同様とし，ラベル付けのなされた仮想店舗内動画 740 本を学習データに用いた．なお，各動画は 1 本あたり 10 枚の連続画像とみなしているため，データセットは計 7400 枚のラベル付き画像となる．また，全 5 種類の環境音と画像との紐付けは，環境音タグにあたる値を CNN アーキテクチャにおける全結合層のユニットに追加することによって行っている（図 3）．

楽曲から予測ラベルを出力する予測器には，SVM を用いた．その際の入力には，各楽曲から算出した MFCC を用いている．

識別器から得られる結果は，各ラベルにおける値である．動画と楽曲のラベルの類似度を計算し，値が大きい 3 曲を推薦する．本研究ではユークリッド距離，コサイン類似度，ピアソ

図 2: 店舗内動画イメージ図

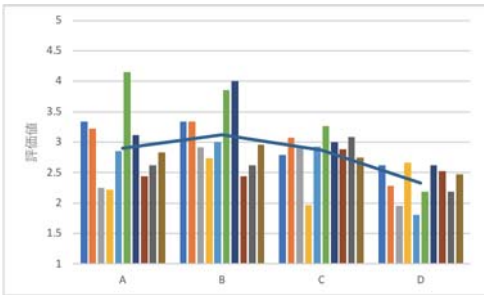

図 4: 楽曲推薦結果

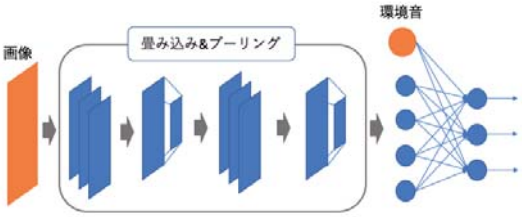

図 3: 店舗識別用 CNN のアーキテクチャ

ン相関係数を用いて求める．

6. 評価実験

提案システムの評価は，曲付きの動画を視聴した評価者に対するアンケートによって行った．評価者は本学学生から募集し，1 本の動画に対し 8 名または 9 名が 5 段階で評価を行った．

提案システムを用いて計 10 本の動画に対する楽曲をそれぞれ 3 曲推薦し，評価した結果を図 4 に示す．A，B，C はそれぞれコサイン類似度，ピアソン相関係数，ユークリッド距離にもとづいて推薦を行った結果であり，D はランダムに推薦を行った結果である．

類似度指標ごとの平均値は 2.99，3.12，2.89 の値で，標準偏差は 0.59，0.46，0.32 であった．また，図 4 からも分かるように A と B においては一部突出して評価が高いものが存在しており，D と比べると平均的にはいずれのグループも評価値が高いことが分かる．

しかし，評価者から感想を聞いたところ，「場所が歌詞と合っていない」「歌詞が気になってしまう」などの意見が得られた．

7. 考察

どの類似度計算においても，定量的にはランダム推薦に勝るものの，定性的には最適な BGM 推薦ができたと結論付けることはできなかった．原因の一つとしてデータの少なさとラベルの多さが学習に影響を与えたと考えられる．また十分なラベル付けがされていないという問題もあった．本研究ではデータセット作成時に動画や楽曲に対してラベル付けをしてもらったが，複数のラベルがつけられたものと，そうでないものの差が激しかった．楽曲に対するタグ付けは 1 曲に対し最低 2 人で行っていたことでラベル付与の偏りが生まれ，推薦に影響が出たのではないかと考えている．

加えて評価実験において同じ楽曲であるにもかかわらず，グループが異なるだけで点数が異なるものが多く存在した．前の曲との関係性が原因と考えられるが，確証を得るためにも人為的に様々な曲と組み合わせ，評価を行う必要がある．

8. おわりに

本稿では，画像と環境音を組み合わせ，店舗ラベルを用いたデータセットを構築することにより，店舗内での雰囲気を反映した BGM 推薦システムを提案した．3 種類の類似度指標を用いて動画に対し楽曲を推薦した結果，いずれの類似度指標においても，ランダムで推薦された曲よりも平均的に高い評価が得られた．一方で，定性的にはあまり高い評価は得られなかった．

今後の課題としては，データセットの拡充や，楽曲間の相性を考慮したシステムの改良などが挙げられる．

参考文献

[1] Daniel Västfjäll. Emotion induction through music: A review of the musical mood induction procedure. *Musicae Scientiae*, Vol. 5, No. 1_suppl, pp. 173–211, 2001.

[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. 情報処理学会研究報告音楽情報科学（MUS）, Vol. 2004, 127（2004-MUS-058）, pp. 33–38, 2004.

[3] Marius Kaminskas and Francesco Ricci. Location-adapted music recommendation using tags. In *International Conference on User Modeling, Adaptation, and Personalization*, pp. 183–194. Springer, 2011.

[4] R Murray Schafer. *The soundscape: Our sonic environment and the tuning of the world*. Simon and Schuster, 1993.

[5] music.usen.com コンシェルジュ, http://music.usen.com（閲覧日：2017 年 12 月 5 日）.

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.