楽曲の風景特徴化に基づく風景アウェア楽曲推薦システム

糸井 勇貴[†] 奥 健太[†] 山西 良典^{††}

† 龍谷大学理工学部 〒 520-2194 滋賀県大津市瀬田大江町横谷 1-5 †† 立命館大学情報理工学部 〒 520-2194 滋賀県草津市野路東1丁目 1-1

E-mail: †t130441@mail.ryukoku.ac.jp, ††okukenta@rins.ryukoku.ac.jp, †††ryama@media.ritsumeikan.ac.jp

あらまし 風景アウェア楽曲推薦システムは、入力された風景をコンテキストして、その風景に合った楽曲を推薦するシステムである。このようなシステムを実現するための課題としては、事前に風景と楽曲とをマッチングさせておくことが必要となる。本研究では、クラウドソーシングにより得られた風景と楽曲とのマッチングデータに基づき、風景を要素とした特徴ベクトルで楽曲を表現する。これを楽曲の風景特徴ベクトルと定義する。この風景特徴ベクトルに基づき、入力風景とマッチする楽曲集合を選択し、その楽曲集合をその入力風景における推薦楽曲集合とする。風景と楽曲とのマッチング精度について、適合率、再現率、F値の観点から四つの提案方式を比較した。結果、風景種別ごとに高精度となった方式が異なったことから、風景種別に応じて方式を変えるハイブリッド方式の検討が必要であることを示唆した。

キーワード 楽曲推薦システム, クラウドソーシング

1. はじめに

自動車は単なる移動手段だけでなく、自動車を運転すること自体が娯楽の一つとなっている。中でも、FM ラジオやカーオーディオなどで音楽を聴きながらドライブを楽しむという人々は多い。特に、最近のカーオーディオや iPod などのポータブル音楽プレーヤでは、大量の楽曲を登録しておくことができ、ドライブ中に自分の好みに合った楽曲を流すことができる。自分の好みだけでなく、爽快な気分の時にはアップテンポな楽曲、落ち着きたい気分のときにはバラード系の楽曲など、そのときの感情によって聴きたい楽曲も変わる。ドライブ時の感情はドライブ中以外の要因も大きく関わっていることは当然であるが、ドライブ時の風景も感情に影響を与える要因の1つと考えられる。例えば、海沿いを走行しているときには爽快な楽曲を、田園風景の中を走っているときには落ち着いた楽曲を聴きたくなるものである。このように、ドライブ時の風景に合わせて楽曲を変えたいという要求がある。

ユーザのそのときのコンテキスト(時間や場所、感情、現在の活動状況など)に合わせた楽曲を再生する推薦システムとして、コンテキストアウェア楽曲推薦システム [1] がある。例えば、位置情報に合った楽曲を推薦するシステム [2] や日時、交通状況、天気などに合った楽曲を推薦するシステム [3] などがある。しかしながら、コンテキストとしてさまざまな要因が挙げられるものの、風景をコンテキストとして扱った楽曲推薦システムはあまり見当たらない。我々が実現を目指す風景アウェア楽曲推薦システムでは、入力された風景をコンテキストして、その風景に合った楽曲を推薦するシステムである。このようなシステムを実現するための課題としては、事前に風景と楽曲とをマッチングさせておくことが必要となる。

風景と楽曲とをマッチングさせるには,風景および楽曲を何らかの形で特徴化し,特徴に基づき風景と楽曲との関係につい

て分析する必要がある.しかしながら,どの風景にどの楽曲がマッチしているかという判断は人間の感性に委ねる部分が大きい.本研究ではクラウドソーシングにより風景と楽曲とのマッチングを行う.クラウドソーシングにより得られた風景と楽曲とのマッチングデータに基づき,風景を要素とした特徴ベクトルで楽曲を表現する.これを楽曲の風景特徴ベクトルと定義する.この風景特徴ベクトルに基づき,入力風景とマッチする楽曲集合を選択し,その楽曲集合をその入力風景における推薦楽曲集合とする.なお本研究では,風景と楽曲とのマッチング方式に焦点をあてる.ドライブ時の風景の識別については本研究の範囲を超える.

本研究の貢献は以下のとおりである.

- クラウドソーシングにより得られた風景と楽曲とのマッチングデータに基づき、楽曲の風景特徴ベクトルを生成した.
- 入力風景と楽曲とのマッチング方式として, top-N 法, 関値ベース法, 類似度ベース法, クラスタリングベース法の四つの方式を提案した.
- 風景と楽曲とのマッチング精度について,適合率,再現率,F値の観点から四つの方式を比較した.結果,風景種別ごとに高精度となった方式が異なったことから,風景種別に応じて方式を変えるハイブリッド方式の検討が必要であることを示唆した.

2. 関連研究

楽曲推薦システムは、楽曲コンテンツ過多の問題に対し、ユーザの楽曲への嗜好(ユーザプロファイル)に基づき、ユーザに合った楽曲を見つけることを支援するためのシステムである [4]. 例えば、Last.fm^(注1) は人気のインターネットラジオである一方で楽曲推薦システムでもある。ユーザがお気に入りとして保

(注1): http://www.last.fm

存した好きな楽曲やアーティストの情報に加え,ユーザの聴取 履歴を基にユーザの楽曲への嗜好を予測することで,楽曲推薦 を行っている.

2.1 コンテキストアウェア楽曲推薦システム

近年では、ユーザのそのときのコンテキスト (時間や場所、感情、現在の活動状況など)を考慮に入れて楽曲を推薦する、コンテキストアウェア楽曲推薦システムも注目されている [1]. これは、ユーザが今どのような楽曲を聴きたいかは、そのときのコンテキストに影響を受けるという考えに基づく.

Reddy ら [3] は、モバイル環境上で、現在のコンテキストに合った楽曲プレイリストを生成する楽曲推薦システム Lifetrakを提案している。コンテキストとして、位置情報や日時、交通状況、天気などが含まれる。与えられたコンテキストにおける楽曲への嗜好評価値を推定し、それを基にプレイリストを生成している。Kaminskas ら [2] は、コンテキストの一つとして位置情報に着目し、位置情報に合った楽曲を推薦する問題に取り組んでいる。タグベースおよび知識ベース手法を組み合わせたハイブリッド型の推薦システムを提案している。本研究では、ドライブ時の楽曲推薦をシチュエーションとして想定しており、ドライブ時に遭遇する風景をコンテキストとして扱っている。

Baltrunas ら [5] は、車内での楽曲推薦を行うシステムとして InCarMusic を提案している。この論文では、ドライビングス タイルや道路風景、交通状況、天気などをコンテキスト因子として扱っている。これらのコンテキスト因子と 10 個の楽曲ジャンル(クラシック、ヒップホップ、Jazz、ポップなど)との関連性について評価値付けの観点から分析している。Baltrunas らの研究と本研究との違いは、本研究では、楽曲と風景とのマッチングデータを基に、楽曲の風景特徴ベクトル化を行い、風景特徴ベクトルに基づく楽曲推薦手法を提案している点である。

2.2 画像に基づく感情アウェア楽曲推薦システム

感情はコンテキストの一つとみなされる. 特に, ユーザの楽曲への嗜好は, そのときの感情に強く影響を受けるといえる. 例えば, 悲しい気分のときと幸せな気分のときとでは, 聴きたい楽曲は大きく変わってくる. 感情アウェア楽曲推薦システムは, ユーザの感情に合った楽曲を推薦する[6].

感情アウェア楽曲推薦を実現するアプローチについては、さまざまな研究が行われているが、画像と楽曲とを対応付けることで、画像を鑑賞しているときの感情と楽曲を鑑賞しているときの感情とをマッチングさせて、楽曲推薦を行う研究もある.

Songhua ら [7] は、言語モデルを用いて画像と楽曲とを対応付けるアプローチを提案している。画像に付与されたアノテーションおよび楽曲の歌詞に言語モデルを適用することで、画像と楽曲とを対応付けている。Stupar ら [8] は、映画から抽出したシーン画像とそのときに流れている楽曲を対とした学習データセットを用意し、楽曲と画像との関係を学習させたモデルを構築している。

本研究では、ドライブ時の楽曲推薦をシチュエーションとして想定し、対象とする画像はドライブ時に遭遇する風景画像として、「湖沿い風景」、「山間風景」、「田園風景」、「川沿い風景」、「海沿い風景」、「市街地風景」に限定した。また、楽曲と画像と

のマッチングデータは、クラウドソーシングにより収集した.

3. 問題設定およびシステム概要

第3.章から第5.章では、提案システムについて説明する.本章では、本研究の問題設定について述べ、提案システムである風景アウェア楽曲推薦システムの概要およびインタフェースについて述べる.

3.1 問題設定

提案システムの目的は,入力された風景にマッチする楽曲を ユーザに推薦することである.まず,風景属性および楽曲集合 の定義を示した上で,問題設定を述べる.

定義:風景属性. 風景属性として, $S=\{s_1:lakeside,s_2:mountain,s_3:paddy,s_4:riverside,s_5:seaside,s_6:urban\}$ の6種類の属性を定義する.それぞれ,湖沿い風景,山間風景,田園風景,川沿い風景,海沿い風景,市街地風景を表す.これらの風景は,Okuらの先行研究 [9] の定義に則り,ドライブ時に遭遇しやすい風景として選択した.

定義:楽曲集合。 推薦対象となる楽曲集合を P と定義する。 i 番目の楽曲を $p_i \in P$ と表す。なお,楽曲を特徴ベクトルとして扱う場合には,明示的に p_i のようにボールド体で表す。 提案システムの問題設定は,与えられた風景属性 $s_j \in S$ にマッチした楽曲 $p_i \in P$ をユーザに推薦することである.

3.2 システム概要

図1は、提案システムの構成図である.以下、各処理手順の概要を説明する.項目の番号は図中の番号に対応する.

- (1) 楽曲データベースに登録されている楽曲集合について, 事前に風景特徴ベクトルを作成しておく. 詳細は第 4. 章で述 べる
- (2) ユーザは提示された風景の中から任意の風景を選択する. 選択された風景を入力風景とし、風景特徴ベクトルに変換する. ここでは、one-hot 表現により、入力風景に対応する要素を 1 とし、それ以外の要素を 0 とした特徴ベクトルを作成する. 例えば、 s_2 : mountain の特徴ベクトルは、 s_2 = (0,1,0,0,0,0) となる.
- (3) 入力風景および楽曲の風景特徴ベクトルに基づき,入力風景と楽曲とをマッチングする.マッチングされた楽曲集合をマッチ楽曲集合とする.詳細は第5.章で述べる.
- (4) 入力風景に対応するマッチ楽曲集合からランダムにk件の楽曲集合を選択し、推薦リストとしてユーザに提示する.

なお、推薦楽曲件数 k および推薦リストの提示方法はアプリケーションに依存する。例えば、ドライブ時の風景にマッチした楽曲を再生するようなアプリケーションの場合は、k=1となりユーザ操作を介さずに推薦楽曲が自動再生される。

4. クラウドソーシングに基づく楽曲の風景特徴 ベクトル化

風景アウェア楽曲推薦を実現するためには、あらかじめ楽曲 集合を風景特徴ベクトル化しておく必要がある. どの楽曲がど の風景にマッチしているかという判断は、人間の感性に委ねる 部分が大きい. 実際にはその感性は個人に依存することも考え

システム構成



(0,0,0,1,0,0)

楽曲DB 」(1)楽曲の風景特徴ベクトル化

 $\frac{1}{2}$ (16.8.1.1.3.1)

オフライン処理

 $\lambda_3(12,1,3,11,1,2)$

 $harphi_{96}(1,0,1,2,1,25)$

★(3)風景と楽曲のマッチング

出力 → 20 → 24 → 47 → 61 → 87 (4)推薦リストの決定

図 1 システム構成図:(1) 楽曲データベース内の各楽曲について風景特徴ベクトル化を行う. (2) 入力風景を特徴ベクトルとして表現しておく。(3) 入力風景の特徴ベクトルと楽曲の風 景特徴ベクトルに基づき、風景と楽曲とをマッチングする。(4)入力風景にマッチングさ れた楽曲集合から k 件の楽曲集合を推薦リストとしてユーザに提示する.

られるが、本研究ではまず一般的な感性の獲得を試み、クラウ ドソーシングを用いる. クラウドソーシングにより得られた楽 曲と風景とのマッチングデータに基づき、楽曲の風景特徴ベク トル化を行う。

4.1 節では、楽曲と風景とのマッチングデータを収集するた めのタスク設定について述べる. 4.2 節では、楽曲と風景との マッチングデータから楽曲の風景特徴ベクトルを作成する方法 について述べる.

4.1 タスク設定

図2に、クラウドソーシングのワーカに提示したタスク画面 のイメージを示す. タスク画面は、タスク部、楽曲部、回答部 から構成される.

タスク部にワーカへのタスク内容を提示している.「聴取した 楽曲と最もマッチしていると感じる風景を、下記に示す六つの 風景の中から一つを選択してください.」と依頼している.

楽曲部には、対象の楽曲ファイル (MP3形式) へのリンクを 提示している. ワーカはリンクをクリックすることで楽曲を聴 取する. 楽曲の聴取範囲は任意としている.

回答部には、6種類の風景写真を提示し、ワーカはこれらの 中から、タスクに合った風景写真を一つ選択する. なお、各風 景写真は、3.1節で定義した風景属性集合 S 内のいずれかの風 景属性に対応する. ここで,順序効果を相殺するため,風景写 真の並び順は毎回ランダムに変更している.

風景写真は ImageNet^(注2) に登録されている写真デー タを参照した、3.1 節で定義した風景属性 S に対応し、 {lakeside, mountain, paddy, riverside, seaside, urban} の各ラ ベルをタグとして付与されている写真データを選択した. 各風



図 2 クラウドソーシングによる楽曲と風景とのマッチングタスクの イメージ. タスク画面は、タスク部、楽曲部、回答部から構成さ れる. ワーカは提示された楽曲について, 六つの風景の中から最 もマッチしていると感じる風景を一つ選択する.

景属性につき、著者の判断により典型的な写真を各30件選択 した. タスク画面へは、該当する風景写真 30 件の中からラン ダムに1件を抽出し提示した.

4.2 楽曲の風景特徴ベクトル化

楽曲 $p_i \in P$ について、風景属性 $s_i \in S$ がマッチしていると 回答したワーカの数を c_{ij} とする. このとき,楽曲風景特徴べ クトル p_i' を次式のように表す.

$$\mathbf{p}_i' = (c_{i1}, c_{i2}, c_{i3}, c_{i4}, c_{i5}, c_{i6}). \tag{1}$$

さらに、ベクトルの要素の和が1となるように正規化したべ

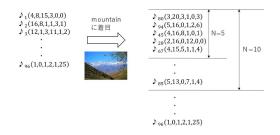


図 3 top-N 法によるマッチング例.

クトルを正規化楽曲風景特徴ベクトル p とし、次式で表す.

$$\mathbf{p}_i = \frac{1}{N}(c_{i1}, c_{i2}, c_{i3}, c_{i4}, c_{i5}, c_{i6}). \tag{2}$$

5. 入力風景と楽曲とのマッチング

本研究では、入力風景 s_j と楽曲 p_i とのマッチング方式として、top-N 法、閾値ベース法、類似度ベース法、クラスタリングベース法の 4 種類の方式を提案する。各方式によりマッチングされた楽曲集合を入力風景 s_j に対するマッチ楽曲集合とする。本節では、各方式について説明する。

5.1 top-N 法

 ${
m top-}N$ 法は多数決により風景と楽曲とをマッチングさせる方式である. ${
m top-}N$ 法では,楽曲風景特徴ベクトル p_i' を用いる.入力風景に対応する要素(s_j において値が 1 である要素)に着目し, p_i' において対応する要素の値が最も高い N 件の楽曲集合をマッチ楽曲集合とする.

図 3 に入力風景を山間風景 s_2 : mountain としたときの例を示す。まず, s_2 : mountain に対応する 2 番目のベクトルの要素に着目する。この要素の値を基準に降順に楽曲集合 P をソートする。ソート後の楽曲集合において上位 N 件の楽曲集合がマッチ楽曲集合となる。図 3 では N=5,10 のときの例を示している。

5.2 閾値ベース法

閾値ベース法は、一定人数により支持されている判断を基に、風景と楽曲とをマッチングさせる方式である。 閾値ベース法では top-N 法と同様、楽曲風景特徴ベクトル p_i' を用いる。 入力風景に対応する要素に着目し、 p_i' において対応する要素の値が 閾値 θ 以上の楽曲集合をマッチ楽曲集合とする。

図 4 に入力風景を山間風景 s_2 : mountain としたときの例を示す. top-N 法と同様に, s_2 : mountain に対応する 2 番目のベクトルの要素に着目する.この要素の値が閾値 θ 以上の楽曲集合がマッチ楽曲集合となる.図 4 では, $\theta=7,9,10,12$ のときの例を示している.

top-N 法および閾値ベース法では,参照元となる楽曲集合は同一となるが,その集合からマッチ楽曲集合を選択する際の基準が異なる. top-N 法がマッチ楽曲集合の件数 N を先に定めてから上位 N 件の楽曲を選択するのに対し,閾値ベース法では閾値以上の楽曲集合をすべて選択する。そのため,閾値ベース法で選択される楽曲件数は参照元の楽曲集合や閾値により変化する.図 3 の top-N 法の N=5 の例では,注目要素の値に

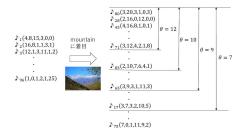


図 4 閾値ベース法によるマッチング例.

依存せず最大 N=5 件の楽曲集合が選択されている.一方で,図 4 の閾値ベース法の $\theta=12$ の例では,注目要素の値が 12 以上の楽曲集合がすべて選択されている.

5.3 類似度ベース法

類似度ベース法は、ベクトル間の類似度に基づき風景と楽曲とをマッチングさせる方式である。類似度ベース法では、正規化特徴ベクトル p_i を用いる。入力風景ベクトル s_j と楽曲の特徴ベクトル p_i とのベクトル間類似度を算出し、類似度が最も高い k 件の楽曲集合をマッチ楽曲集合とする。

本方式ではベクトル間類似度の算出にコサイン類似度を用いる. ベクトル $x \ge y$ のコサイン類似度は次式で算出する.

$$cos(\boldsymbol{x}, \boldsymbol{y}) = \frac{\sum_{i} x_{i} y_{i}}{\sqrt{\sum_{i} x_{i}^{2}} \sqrt{\sum_{i} y_{i}^{2}}}.$$
 (3)

 $cos(m{s}_j,m{p}_i)$ が最も高い k 件の楽曲集合をマッチ楽曲集合とする.

 ${
m top-}N$ 法および閾値ベース法が入力風景に対応する要素のみに着目するのに対し,類似度ベース法は全要素を基に類似度を算出する.例えば,次のような入力風景ベクトル ${m s_i}$ および二つの楽曲風景特徴ベクトル ${m p_1'}, {m p_2'}$ を考える.

$$\mathbf{s}_{i}' = (1, 0, 0, 0, 0, 0), \tag{4}$$

$$\mathbf{p}_1' = (15, 15, 0, 0, 0, 0), \tag{5}$$

$$\mathbf{p}_2' = (15, 3, 3, 3, 3, 3). \tag{6}$$

このとき,top-N 法および閾値ベース法では,入力風景に対応する第一要素のみに着目するため, p_1' も p_2' も同順位の楽曲データとしてみなされる.一方で,類似度ベース法では,ベクトル間の類似度がそれぞれ, $cos(s_j,p_1)=0.707,cos(s_j,p_2)=0.913$ と算出され—厳密には正規化特徴ベクトルに基づき算出される—, p_2 の方が順位が高くなる.これは,入力風景ベクトルの分布により近い方が優先的に候補楽曲として取り出されることを意味する.

5.4 クラスタリングベース法

クラスタリングベース法は、あらかじめ楽曲集合をクラスタリングしておき、クラスタ単位で風景とマッチングさせる方式である。クラスタリングベース法では、正規化特徴ベクトル p_i を用いる。あらかじめ楽曲集合Pを p_i に基づきm個のクラスタにクラスタリングしておく。その後、入力風景ベクトル p_i に最近傍のクラスタに属する楽曲集合をマッチ楽曲集合とする。本方式では、クラスタリング手法としてk-means 法を用いる。

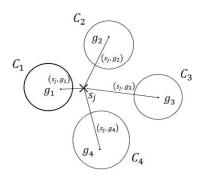


図 5 クラスタリング法によるマッチング例.

以下,クラスタリングベース法による入力風景と楽曲とのマッチング方法について説明する.

- (1) 正規化特徴ベクトル p_i に基づき、楽曲集合 P を m 個のクラスタ C_1, C_2, \ldots, C_m に分類する.
 - (2) 各クラスタの重心 $g_1, g_2, ..., g_m$ を算出する.
- (3) 入力風景ベクトル s_j と各クラスタの重心との距離を算出する。距離の算出にはユークリッド距離を用いる。ベクトル x と y のユークリッド距離は次式で算出する。

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i} (x_i - y_i)^2}.$$
 (7)

- (4) $d(\mathbf{s}_i, \mathbf{g}_l)$ が最小となるクラスタ C_l を選択する.
- (5) クラスタ C_l に属する楽曲集合をマッチ楽曲集合とする.

図 5 にクラスタ数を m=4 としたときの例を示す.図 5 では,入力風景 \mathbf{s}_j の最近傍クラスタとしてクラスタ C_1 が選択されている.このクラスタ C_1 に属する楽曲集合がマッチ楽曲集合となる.

6. マッチング方式の比較評価

第 5. 章で提案した 4 種類のマッチング方式の違いについて 比較評価する。6.1 節では,評価実験に用いるデータセットに ついて述べる。6.2 節では,比較手法について述べる。6.3 節 では,評価実験の方法,また比較の際に用いる,実験から算出 される数値について述べる。最後に 6.4 節では,算出された数 値から考察されることを述べる。

6.1 データセット

楽曲データセットとして、『クラシック名曲サウンドライブラリー』 $^{^{({
m lt}3)}}$ のクラシック楽曲を利用した、楽曲利用については当サイト作成者の許諾を得ている。本実験では、当サイトの楽曲データの中から、管弦楽曲 96 曲(|P|=96)を利用した。

この楽曲データセットを対象に、4.1 で説明したクラウドソーシングによるタスクを依頼した。1 曲あたり 30 名(N=30)のワーカにタスクを依頼した。結果、2,880 件の楽曲-風景マッチングデータを得た。

6.2 比較方式

第 5. 章で説明した, top-N 法, 閾値ベース法, 類似度ベース法, クラスタリングベース法の四つのマッチング方式を比較方式とする. 各手法による推薦リストおよびパラメタの設定は以下のとおりである.

a) top-N 法 (top5, top10)

N=5,10 としたとき, top-N 法によるマッチ楽曲集合を得る.

b) 閾値ベース法(θ 7, θ 9, θ 10, θ 12)

 $\theta = 7,9,10,12$ としたとき,閾値ベース法によるマッチ楽曲集合を得る.

c) 類似度ベース法 (cos10)

コサイン類似度が最も高い 10 件の楽曲集合をマッチ楽曲集合とする.

d) クラスタリングベース法(cl5, cl7, cl10, cl12, cl15) m=5,7,10,12,15 としたとき,クラスタリングベース法によるマッチ楽曲集合を得る.

6.3 評価方法

まず,実験手順は次のとおりである.

- (1) データセットを無作為に 5 組に分割し, 4 組を学習データ, 残り 1 組をテストデータとする.
- (2) 学習データに基づき、各楽曲の風景特徴ベクトルを作成する.
- (3) 入力風景 $s_1, s_2, ..., s_6$ ごとに,テストデータに含まれる各楽曲について,各方式によるマッチ楽曲集合を得る.
- (4) マッチ楽曲集合とテストデータの楽曲に元々付与されている風景属性ラベルとを比較し、一致していれば適合と判定する.この判定に基づき、方式ごとに適合率、再現率、F値を算出する.
- (5) 学習データとテストデータの組合せを変えて, (1) から (4) の手順を 5 回繰り返し, その平均値を取得する.

適合率は正確性に関する指標であり、システムにより出力されたデータ集合のうち、正解であるデータ集合の割合を指す、 再現率は網羅性に関する指標であり、正解である全データ集合のうち、システムにより出力されたデータ集合の割合を指す.

正解かつシステムにより出力されたデータ集合を TP, 正解かつシステムにより出力されなかったデータ集合を TN, 不正解かつシステムにより出力されたデータ集合を FP, 不正解かつシステムにより出力されなかったデータ集合を FN とすると、適合率 precision および再現率 recall は次式で求められる.

$$precision = \frac{|TP|}{|TP| + |FP|},\tag{8}$$

$$recall = \frac{|TP|}{|TP| + |TN|}. (9)$$

さらに、正確性と網羅性を総合的に評価するための指標として、F値がある。F値 F-measure は次式で求められる。F値が高いシステムほど良いシステムと判断される。

$$F\text{-}measure = \frac{2 \cdot precision \cdot recall}{precision + recall}.$$
 (10)

表 1 入力風景ごとの各方式の適合率. ボールド体は各風景における 最高値を示す.

	lakeside	mountain	paddy	riverside	seaside	urban	average
top5	0.350	0.439	0.292	0.496	0.245	0.531	0.392
top10	0.362	0.437	0.255	0.463	0.196	0.583	0.382
$\theta 7$	0.290	0.324	0.247	0.317	0.215	0.368	0.294
θ 9	0.315	0.343	0.352	0.360	0.229	0.425	0.338
$\theta 10$	0.358	0.396	0.355	0.404	0.253	0.498	0.377
$\theta 12$	0.347	0.387	0.524	0.451	0.111	0.551	0.395
$\cos 10$	0.332	0.424	0.246	0.463	0.200	0.596	0.377
cl5	0.265	0.364	0.205	0.410	0.063	0.528	0.306
cl7	0.267	0.384	0.229	0.427	0.148	0.567	0.337
cl10	0.339	0.367	0.269	0.438	0.164	0.547	0.354
cl12	0.349	0.423	0.252	0.456	0.251	0.575	0.384
cl15	0.302	0.403	0.275	0.494	0.252	0.545	0.378
average	0.323	0.391	0.292	0.432	0.194	0.526	

表 2 入力風景ごとの各方式の再現率. ボールド体は各風景における 最高値を示す.

	lakeside	mountain	paddy	riverside	seaside	urban	average
top5	0.102	0.101	0.180	0.145	0.217	0.128	0.146
top10	0.217	0.181	0.257	0.246	0.353	0.274	0.255
θ 7	0.479	0.664	0.279	0.521	0.202	0.632	0.463
$\theta 9$	0.307	0.515	0.136	0.434	0.130	0.460	0.330
$\theta 10$	0.194	0.340	0.097	0.329	0.078	0.428	0.244
$\theta 12$	0.147	0.249	0.059	0.258	0.028	0.400	0.190
$\cos 10$	0.185	0.171	0.279	0.252	0.359	0.277	0.254
cl5	0.406	0.310	0.352	0.337	0.251	0.423	0.346
cl7	0.254	0.175	0.273	0.294	0.201	0.371	0.261
cl10	0.192	0.086	0.245	0.239	0.185	0.272	0.203
cl12	0.185	0.088	0.188	0.172	0.160	0.229	0.170
cl15	0.114	0.056	0.156	0.142	0.160	0.133	0.127
average	0.232	0.245	0.208	0.281	0.194	0.336	

表 3 入力風景ごとの各方式の F 値. ボールド体は各風景における最高値を示す.

	lakeside	mountain	paddy	${\bf riverside}$	seaside	urban	average
top5	0.158	0.164	0.223	0.224	0.230	0.206	0.201
top10	0.271	0.256	0.256	0.322	0.252	0.373	0.288
θ 7	0.361	0.436	0.262	0.394	0.208	0.465	0.355
θ 9	0.311	0.412	0.196	0.394	0.166	0.442	0.320
$\theta 10$	0.251	0.366	0.152	0.363	0.119	0.460	0.285
$\theta 12$	0.206	0.303	0.106	0.328	0.045	0.464	0.242
$\cos 10$	0.237	0.244	0.261	0.326	0.257	0.379	0.284
cl5	0.321	0.335	0.259	0.370	0.101	0.470	0.309
cl7	0.260	0.241	0.249	0.348	0.171	0.448	0.286
cl10	0.245	0.139	0.257	0.309	0.174	0.363	0.248
cl12	0.242	0.146	0.216	0.250	0.195	0.327	0.229
cl15	0.166	0.099	0.199	0.221	0.196	0.214	0.182
average	0.253	0.262	0.220	0.321	0.176	0.384	

6.4 結 果

表 1,表 2,表 3 は、それぞれ入力風景ごとの各方式の適合率、再現率、F値である。各風景において最高値を示した手法については、その数値をボールド体で示している。

a) クラスタリングベース法の比較

表 3 のクラスタリングベース法 (cl5, cl7, cl10, cl12, cl15) の平均値に着目すると,クラスタ数 m が小さいほど F 値が高

いことがわかる.

b) 閾値ベース法の比較

表 3 の閾値ベース法 $(\theta 7, \theta 9, \theta 10, \theta 12)$ の平均値に着目する と,閾値 θ が小さいほど F 値が高いことがわかる.表 2 および 表 1 からわかるように,当然閾値を下げた方が再現率は向上するが適合率は低下する.ただし,適合率の低下率に比べ,再現率の向上率が高いため,F 値の観点からは閾値を下げた方が精度が高くなったといえる.

c) 風景種別間の比較

表3について風景種別ごとの平均値を比較すると、urban が最も高い F 値を示したことがわかる。他の5種類の風景が自然風景であるのに対し、urban は市街地風景ということで唯一人工的な風景であった。クラウドソーシングにおいて楽曲にマッチする風景を選択する際、lakeside や seaside のように類似する自然風景に対しては、判断に揺らぎが生じる。一方で、楽曲が人工的な風景にマッチすると判断する際には、確定的にurban が選択される。このことが urban における F 値が最高値を示した原因であるといえる。

d) 風景種別ごとの方式の優位性の比較

いずれの表を参照しても,風景種別によって最高値を示した方式が異なることがわかる.F値に着目すると,表 3 より,lakeside,mountain,paddy,riverside については θ 7 が最高値を示した.一方で,seaside については $\cos 10$ が,urban については $\cos 10$ が最も高かった.このように,風景種別によって適切な方式が異なることから,今後は風景種別に応じた方式選択を行うハイブリッド方式の導入を検討する必要がある.

7. 結 論

本研究では風景アウェア楽曲推薦システムの実現を目指し、クラウドソーシングにより得られた風景と楽曲とのマッチングデータに基づき、楽曲の風景特徴ベクトルを生成した. さらに、入力風景と楽曲とのマッチング方式として、top-N法、関値ベース法、類似度ベース法、クラスタリングベース法の四つの方式を提案した.

風景と楽曲とのマッチング精度について,適合率,再現率, F値の観点から四つの方式を比較した.結果,風景種別によっ て適切な方式が異なることから,今後は風景種別に応じた方式 選択を行うハイブリッド方式の導入を検討する必要がある.

今後の展望として,我々が別途研究課題として進めている道路風景の識別成果[9]と組み合わせて,風景アウェアドライブ楽曲推薦システムの研究に取り組む.

謝辞

本研究は JSPS 科研費 15K12151 の助成を受けたものです. ここに記して謝意を表します.また,本研究で利用した楽曲データセットは,『クラシック名曲サウンドライブラリー』(http://classical-sound.seesaa.net/)のサイト作成者のご厚意により使用許諾を頂きました.ここに謝意を表します.

抽 女

[1] Marius Kaminskas and Francesco Ricci. Contextual music

- information retrieval and recommendation: State of the art and challenges. *Computer Science Review*, Vol. 6, No. 2-3, pp. 89–119, may 2012.
- [2] Marius Kaminskas, Francesco Ricci, and Markus Schedl. Location-aware music recommendation using auto-tagging and hybrid matching. In *Proceedings of the 7th ACM con*ference on Recommender systems, pp. 17–24, New York, New York, USA, 2013. ACM Press.
- [3] S. Reddy and Jeff Mascia. Lifetrak: Music In Tune With Your Life Categories and Subject Descriptors. In Proceedings of the 1st ACM International Workshop on Humancentered Multimedia, pp. 25–34, 2006.
- [4] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor. Recommender Systems Handbook. Springer, 2011.
- [5] Linas Baltrunas, Marius Kaminskas, Bernd Ludwig, Omar Moling, Francesco Ricci, and Aykan Aydin. InCarMusic: Context-Aware Music Recommendations in a Car. In Proceedings of the 12th International Conference on Electronic Commerce and Web Technologies, pp. 89–100, 2011.
- [6] Yong Zheng, Bamshad Mobasher, and Robin D Burke. The role of emotions in context-aware recommendation. In Proceedings of the RecSys 2013 Workshop on Human Decision Making in Recommender Systems, pp. 21–28, 2013.
- [7] Xu Songhua, Jin Tao, and C. M Lau Francis. Automatic generation of music slide show using personal photos. In Proceedings of the 10th IEEE International Symposium on Multimedia, pp. 214–219, 2008.
- [8] Aleksandar Stupar and Sebastian Michel. PICASSO To Sing you must Close Your Eyes and Draw. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information, pp. 715–724, 2011.
- [9] Kenta Oku, Ryosuke Yamanishi, Kohei Matsumura, and Kyoji Kawagoe. Drive scenery estimation using image features extracted from roadmap and satellite images. In Proceedings of the 16th International Symposium on Advanced Intelligent Systems, pp. 1575–1589, 2015.