

景観アウェア楽曲推薦の個人化のための共クラスタモデルの検討

甲田翔太郎[†] 奥 健太[†]

[†] 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
E-mail: †t140458@mail.ryukoku.ac.jp, ††okukenta@rins.ryukoku.ac.jp

あらまし 我々は、景観に合った楽曲を推薦する景観アウェア楽曲推薦システムの研究に取り組んでいる。これまでにクラウドソーシングにより収集した楽曲-景観関係データに基づいた楽曲推薦システムを提案し、実装してきた。楽曲-景観関係データは、ワークに楽曲を提示し、その楽曲に最もマッチしているという観点で景観を選択してもらうことで作成している。しかしながら、先行研究においては、どの楽曲がどの景観にマッチしているかという点に関しては個人差が大きいことを確認した。そこで、本研究では、楽曲推薦の個人化に向けて楽曲-景観関係データに対し、共クラスタリングを適用することで景観ごとのユーザと楽曲の共クラスタモデルを獲得する。具体的には、共クラスタリング手法の一つである確率的ブロックモデルにより楽曲-景観関係データの共クラスタモデルを得る。得られた共クラスタモデルにおいて、各クラスタにおける楽曲の特徴について分析する。

キーワード 楽曲推薦, 共クラスタリング, 確率的ブロックモデル

1. はじめに

ドライブを趣味としている人は多くいるが、ドライブの魅力として、景観や楽曲を楽しみながら運転することを楽しむ人は多くいる。近年はカーオーディオが発展しており、事前に楽曲を登録しておけばドライブ中に自分の好みにあった楽曲を流すことができる。しかし、人の好みは一定ではなくその時の感情によって常に変動するものである。ドライブ時の感情に影響を与える要素としては様々なものが考えられるが、そのうちのひとつとして、景観が考えられる。例えば、のどかな田園地帯を走っているときは静かな楽曲を、大都会を走っているときは明るい楽曲を聴きたくなるものである。このように、景観によって聴きたい楽曲は異なる。

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、景観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景観アウェア楽曲推薦システムでは、入力された景観をコンテキストとして、その景観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、事前に景観ごとにユーザの嗜好と楽曲をマッチングさせておくことが必要となる。しかしながら、どの景観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多い。本研究ではクラウドソーシングにより景観ごとのユーザ-楽曲関係データの収集を行う。クラウドソーシングにより得られた景観ごとのユーザ-楽曲関係データに基づき、景観ごとの共クラスタモデルを作成する。このクラスタモデルに基づき、入力景観とマッチする楽曲集合を選択し、その楽曲集

合に属する楽曲がユーザに推薦される楽曲の対象となる。なお本研究では、個人の嗜好に応じた楽曲推薦を実現するための共クラスタモデルの作成に焦点をあてる。ドライブ時の景観の識別については本研究の範囲を超える。

本稿では、ユーザ-楽曲関係データに対し共クラスタリングを行った結果について定性的に分析する。景観要素ごとに共クラスタリングにより獲得されたクラスタのうち、代表的なクラスタに注目し、そのクラスタに属する楽曲の傾向について分析する。その結果を踏まえながら、景観アウェア楽曲推薦の個人化方法について議論する。

2. 関連研究

2.1 楽曲推薦システム

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 [1]。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。

2.2 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）を基に楽曲を推薦する、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 [2]。これは、そのときのユーザのコンテキストによってユーザが聴きたい楽曲は変移す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

(注1) : <http://www.last.fm>

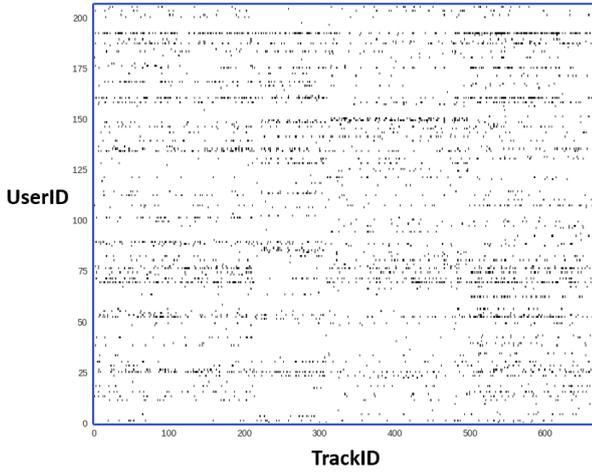


図1 山間風景のマッチングデータ.

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 [4] 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、個人の嗜好に対応した楽曲推薦の実現に向けて、楽曲と景観とのマッチングデータを基に、共クラスタモデルを作成している点である。

2.3 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、悲しい気分のとときと幸せな気分のとときとでは、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、「山林風景」、「田園風景」、「水辺風景」、「都市風景」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。図 1 に山間風景のマッチングデータを示す。図中の縦軸はユーザ、横軸は楽曲を表している。また、青い要素はユーザが風景と楽曲がマッチしていると回答したことを表している。

3. 問題定義

定義 1: 景観要素. 奥らの先行研究 [9] では、予備実験により、道路景観を構成する基本要素として、田園系 (r)、山林系 (m)、水辺系 (w)、都市系 (u) の四つの要素を定義した。これらの要素を景観要素とよぶ。

定義 2: 景観要素 k のユーザ-楽曲関係データ. 景観要素 k のユーザ-楽曲関係データとは、景観要素 k においてユーザ集合 U と楽曲集合 T の関係を表したデータであり、ユーザ-楽曲関係行列として表す。ユーザ-楽曲関係行列 $R^k = [r_{ij}]_{|U| \times |T|}$ は、 $|U| \times |T|$ の行列であり、行はユーザ $u_i \in U$ を、列は楽曲 $t_j \in T$ を表す。行列の要素 $r_{ij}^k \in \{0, 1\}$ は u_i と t_j の関係を表す。ユーザ u_i が楽曲 t_j に対し景観要素 k にマッチするという印象をもつというデータが観測されたとき $r_{ij}^k = 1$ となり、未観測であるとき $r_{ij}^k = 0$ となる。定義 1 に示したとおり、本研究では景観要素として四つの要素を定義しているため、得られるユーザ-楽曲関係行列は R^r, R^m, R^w, R^u —順に田園系、山林系、水辺系、都市系の関係行列—の四つとなる。

4. ユーザ-楽曲関係データの収集

3. 章で定義したとおり、ユーザ-楽曲関係データはユーザ集合 U と楽曲集合 T との関係を表したデータであり、景観要素ごとに用意する。このような関係データを作成するためには、ユーザ u_i にとって楽曲 t_i が景観要素 k にマッチしているかについて知る必要がある。本研究では、まずクラウドソーシングによりユーザ-楽曲-景観関係データを収集し、このデータを基に景観要素ごとのユーザ-楽曲関係データを得る。

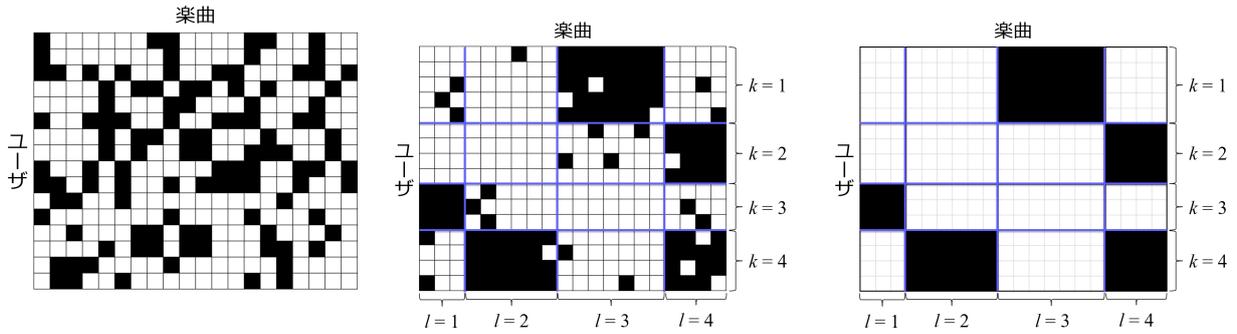
任意の楽曲には楽曲ファイル (MP3 形式, WAV 形式など) が含まれる。事前に全楽曲について特定の区間を切り出し、評価用楽曲ファイルとしておく。楽曲 $t_i \in T$ について、以下の手順によりクラウドソーシングを用いてユーザ-楽曲-景観関係データを収集する:

(1) クラウドソーシングの作業員 (以降、ワーカ) は評価用楽曲ファイルを聴取する。

(2) ワーカは聴取した楽曲と最もマッチしていると感じる景観イメージとして、「田園景観」、「山林景観」、「水辺景観」、「都市景観」、「いずれにも該当しない」の選択肢の中から一つを選択する。何らかの不具合により楽曲が再生されなかった場合は「楽曲が再生されない」を選択する。

1 楽曲あたり複数名のワーカが回答する。1 名のワーカによる 1 楽曲に対する回答を 1 タプルとして

`workers-tracks-landscape` テーブルに登録する。このテーブ



(a) ユーザ-楽曲行列.

(b) 共クラスタ.

(c) 抽象化共クラスタ.

図 2 ユーザ-楽曲行列への共クラスタリングの適用. (a) ユーザ-楽曲行列 R^r を可視化したものである. R^r は $|U| \times |T|$ の行列であり, 行はユーザ $u_i \in U$ を, 列は楽曲 $t_j \in T$ を表す. $r_{ij}^r \in R^r$ は u_i, t_j に対応する要素を表す. 黒い要素は $r_{ij}^r = 1$ を, 白い要素は $r_{ij}^r = 0$ の要素を表す. (b) ユーザ-楽曲行列 R^r に共クラスタリングを適用した結果. クラスタ数を 4×4 としている. $k \in \{1, 2, 3, 4\}$ が, t_j には楽曲クラスタ番号 $l \in \{1, 2, 3, 4\}$ は, それぞれユーザ側, 楽曲側のクラスタ番号を表す. (b) では, R^r をユーザ側, 楽曲側両方のクラスタ番号でソート済みのものを示している. (c) (b) を抽象化したものである. 要素数に対する黒い要素の割合が一定の閾値以上のブロックを黒く, それ以外を白く表示している. これら黒いブロックがユーザと楽曲の関係が強いクラスタを表す.

ルは次の関係スキーマで定義される:

```
workers-tracks-landscape(worker_id,
    track_id, landscape)
```

ここで, `worker_id` はワーカー ID (ユーザ ID), `track_id` は楽曲 ID, `landscape` は選択された景観要素をそれぞれ示す.

`workers-tracks-landscape` テーブルを基に, 景観要素ごとのユーザ-楽曲行列 R^r, R^m, R^w, R^u を作成する. 例えば, `workers-tracks-landscape` テーブルにタプル $(u_i, t_j, rural)$ が存在するとき $r^{r_{ij}} = 1$ となり, そのタプルが存在しないとき $r^{r_{ij}} = 0$ となる. R^m, R^w, R^u についても同様である.

5. ユーザ-楽曲行列の共クラスタリング

得られたユーザ-楽曲行列 R^r, R^m, R^w, R^u に対し, 共クラスタリングを適用する. 本研究では共クラスタリング手法として確率的ブロックモデルを用いる. ここでは, $R^r = [r_{ij}^r]_{|U| \times |T|}$ について説明する.

例えば, 図 2(a) のようなユーザ-楽曲行列 R^r が与えられたとする. R^r は $|U| \times |T|$ の行列であり, 行はユーザ $u_i \in U$ を, 列は楽曲 $t_j \in T$ を表す. $r_{ij}^r \in R^r$ は u_i, t_j に対応する要素を表す. 図中の黒い要素は $r_{ij}^r = 1$ を, 白い要素は $r_{ij}^r = 0$ の要素を表す.

この R^r に共クラスタリングを適用すると, 図 2(b) のようにクラスタリングされる. 図 2(b) ではクラスタ数を 4×4 としている. クラスタリングを実行することで, u_i にはユーザクラスタ番号 $k \in \{1, 2, 3, 4\}$ が, t_j には楽曲クラスタ番号 $l \in \{1, 2, 3, 4\}$ がそれぞれ付与される. 図 2(b) は R^r をユーザ側, 楽曲側両方のクラスタ番号でソート済みのものである. 例えば, $(k, l) = (1, 3)$ や $(k, l) = (4, 2)$ においては黒い要素が集

まったブロックが形成されていることがわかる.

図 2(c) は図 2(b) を抽象化したものである. 要素数に対する黒い要素の割合が一定の閾値以上のブロックを黒く, それ以外を白く表示している. 例えば, 先の $(k, l) = (1, 3)$ および $(k, l) = (4, 2)$ は黒いブロックとして表示している. これら黒いブロックがユーザと楽曲の関係が強いクラスタを表す. このクラスタ内のユーザは田園景観において同じ楽曲を選択していることから, 感性が類似しているといえる. このような共クラスタリングを景観ごとのユーザ-楽曲行列に適用する.

6. 定性的分析

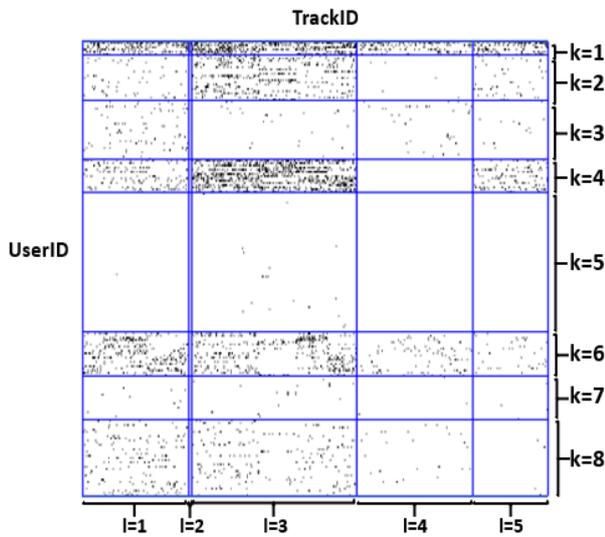
ユーザ-楽曲行列に対し共クラスタリングを行った結果について定性的に分析する. 6.1 節では, 本分析に用いるデータセットについて述べる. 6.2 節では, 景観要素ごとに共クラスタモデルについて分析する. 具体的には, 景観要素ごとに代表的なクラスタに注目し, そのクラスタに属する楽曲の傾向について分析する. 6.3 節では, 6.2 節の結果を踏まえながら, 景観ウェア楽曲推薦の個人化方法について議論する.

6.1 データセット

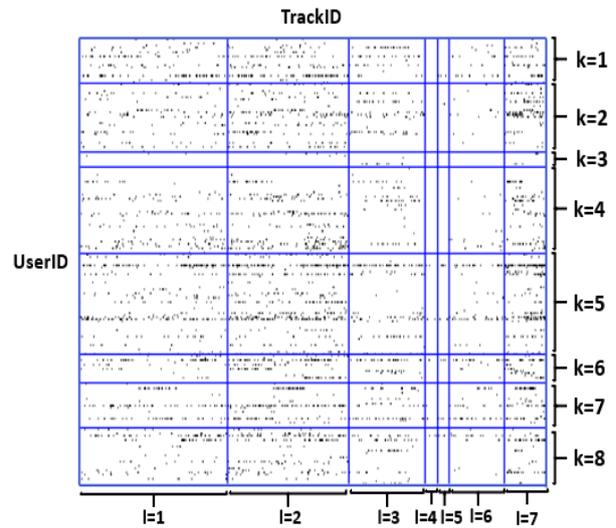
楽曲データセットとして、『クラシック名曲サウンドライブラリー』^(注2)のクラシック楽曲の 663 曲を利用した. 楽曲の研究利用については当サイト作成者の許諾を得ている.

この 663 曲の楽曲について, 4. 章で述べた方法により, ユーザ-楽曲-景観関係テーブル `workers-tracks-landscape` を作成した. クラウドソーシングにおける評価用ファイルとして, 各楽曲ファイルから楽曲長の中心から前後 30 秒間の区間を抽出した. ワーカーには 1 楽曲あたり抽出した 30 秒間すべてを

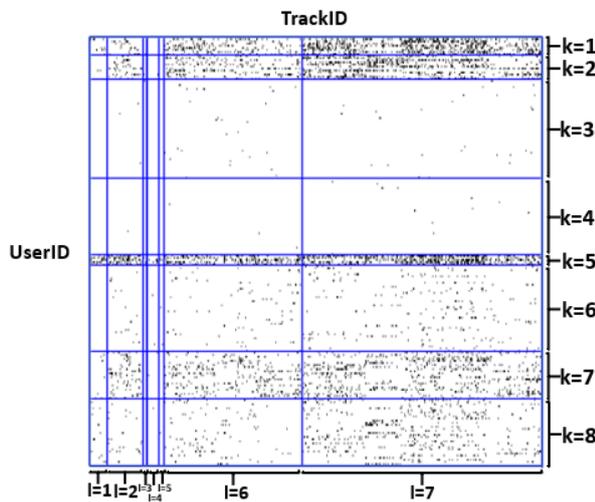
(注2): <http://classical-sound.seesaa.net/>



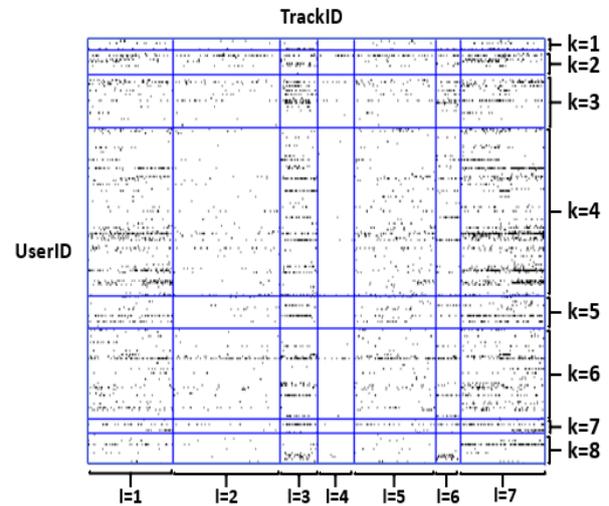
(a) 田園景観.



(b) 山林景観.



(c) 水辺景観.



(d) 都市景観.

図 3 各景観要素に対する共クラスタモデル. 各モデルは行がユーザを, 列が楽曲を表す. 図中の黒い要素は $r_{ij}^r = 1$ を, 白い要素は $r_{ij}^r = 0$ の要素を表す. k, l はそれぞれユーザ, 楽曲のクラスタ番号を示す. それぞれ, R^k をユーザ側, 楽曲側両方のクラスタ番号でソート済みのものである.

表 1 景観要素ごとの評価値数および密度. 密度は 評価値数 / (208 × 663) で算出される.

景観要素	評価値数	密度
rural	4,056	0.029
mountain	3,824	0.028
water	4,957	0.036
urban	5,284	0.038
その他	1,765	-
エラー	4	-

聴取してもらった. 1 楽曲あたり 30 名のワーカに回答してもらった. したがって, $663 \times 30 = 19,890$ のタプルからなる `workers-tracks-landscape` テーブルを作成した. クラウドソーシングに参加したユニークなユーザ数は 208 名であった.

得られた `workers-tracks-landscape` テーブルを基に, 各景観要素のユーザ-楽曲関係行列 R^r, R^m, R^w, R^u を作成した. 各関係行列は 208×663 の行列となる. 表 1 に景観要素ごとの評価値数および密度を示す. ここで, 密度は 評価値数 / (208 × 663) で算出される.

6.2 結果

図 3 に各ユーザ-楽曲関係行列に対し共クラスタリングを適

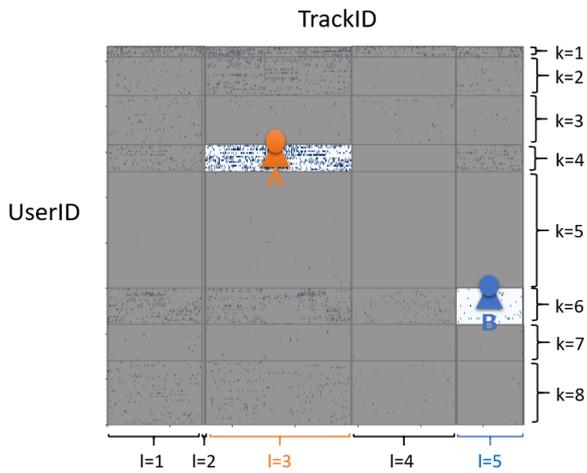


図4 共クラスタに基づく楽曲推薦のイメージ。

用した結果を示す。それぞれ、(a) 田園景観、(b) 山林景観、(c) 水辺景観、(d) 都市景観における共クラスタモデルを表す。各モデルは行がユーザを、列が楽曲を表す。図中の黒い要素は $r_{ij}^k = 1$ を、白い要素は $r_{ij}^k = 0$ の要素を表す。図中の k, l はそれぞれユーザ、楽曲のクラスタ番号を示す。図3は R^k をユーザ側、楽曲側両方のクラスタ番号でソート済みのものである。

以下、景観要素ごとに、代表的なクラスタに注目し、そのクラスタに属する楽曲の傾向について分析する。

a) 田園景観 (表3(a))

$(k, l) = (4, 3)$ のクラスタに属する楽曲を聴取してみると、比較的メロディが高音であり、弦楽器を用いている楽曲が多い傾向がみられた。一方、 $(k, l) = (6, 1)$ のクラスタに属する楽曲には、比較的メロディが高音であり、ピアノを用いている楽曲が多い傾向がみられた。

b) 山林景観 (表3(b))

$(k, l) = (5, 2)$ のクラスタに属する楽曲には、比較的メロディが高音であり、ピアノを用いた楽曲が多いという傾向がみられた。一方、 $(k, l) = (5, 1)$ のクラスタに属する楽曲には、比較的落ち着いたテンポ、協奏曲、ピアノを用いる楽曲が多いという傾向がみられた。

c) 水辺景観 (表3(c))

$(k, l) = (1, 7)$ のクラスタに属する楽曲には、ピアノを用いた楽曲、交響曲の割合が高かった。また、比較的メロディが高音であり、静かなイメージの楽曲が多いという傾向がみられた。一方、 $(k, l) = (5, 6)$ のクラスタに属する楽曲には、のどかな雰囲気、交響曲、管弦楽曲が多いという傾向がみられた。

d) 都市景観 (表3(d))

$(k, l) = (4, 7)$ のクラスタに属する楽曲には、比較的メロディが高音であり、交響曲が多いという傾向がみられた。一方、 $(k, l) = (4, 1)$ のクラスタに属する楽曲には、明るいイメージの楽曲、交響曲が多いという傾向がみられた。

6.3 議論

前節の結果より、同じ景観要素であっても、クラスタによって属する楽曲の曲調の傾向が異なることがわかる。例えば、田園景観においては、ユーザクラスタ $k = 4$ のユーザ集合と $k = 6$

のユーザ集合では田園景観にマッチしていると感じる楽曲の傾向が異なる。前者は、田園景観であれば、メロディが高音であり、弦楽器を用いている楽曲がマッチしていると感じるのに対し、後者は、メロディが高音であり、ピアノを用いている楽曲がマッチしていると感じる。

このような関係を事前に共クラスタモデルとして獲得しておくことで、将来的に景観アウェア楽曲推薦の個人化につなげることができる。図6.1を例に説明する。いくつかのサンプルに基づき、田園景観においてはユーザAがクラスタ $(k, l) = (4, 3)$ に属することがわかっていたとする。このとき、ユーザAが田園景観というコンテキストにおかれている場合には、クラスタ $(k, l) = (4, 3)$ に属する楽曲を推薦すれば良い。同様に、ユーザBがクラスタ $(k, l) = (6, 5)$ に属することがわかっている場合、ユーザBには田園景観において、クラスタ $(k, l) = (6, 5)$ に属する楽曲を推薦すれば良い。

ここで、推薦対象となるユーザが新規ユーザの場合、あらかじめ属するクラスタが明らかでない。このような状況においては、適合性フィードバックなど、システムとの数回のインタラクションを通して徐々に属するクラスタを推定していくなどの対応が必要となる。この所属クラスタの推定は別途検討課題とする。

7. おわりに

本研究では、景観に合った楽曲を推薦する景観アウェア楽曲推薦システムの個人化のための共クラスタモデルを検討した。クラウドソーシングにより得られた景観ごとのユーザ-楽曲関係データに基づき、景観ごとの共クラスタモデルを作成した。作成された共クラスタモデルについて定性的に分析した。具体的には、景観要素ごとに共クラスタリングにより獲得されたクラスタのうち、代表的なクラスタに注目し、そのクラスタに属する楽曲の傾向について分析した。

分析の結果、同じ景観要素であっても、クラスタによって属する楽曲の曲調の傾向が異なることがわかった。このような関係を事前に共クラスタモデルとして獲得しておくことで、将来的に景観アウェア楽曲推薦の個人化につなげることができる点を議論した。

今後は、本稿で得られた知見を基に景観アウェア楽曲推薦システムの個人化手法の具体化に向けた検討を行う。具体的には、推薦対象となるユーザが新規ユーザの場合でも個人化推薦が可能となるように、適合性フィードバックなど、システムとの数回のインタラクションを通して徐々に属するクラスタを推定していくなどの手法を検討する。

謝辞

本研究はJSPS科研費JP15K12151の助成を受けたものです。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。ここに記して謝意を表します。

文 献

- [1] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor. *Recommender Systems Handbook*. Springer, 2011.
- [2] Marius Kaminskis 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.
- [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 Human-centered Multimedia*, pp. 25–34, 2006.
- [4] Marius Kaminskis, Francesco Ricci, and Markus Schedl. Location-aware music recommendation using auto-tagging and hybrid matching. In *[RecSys2013]Proceedings of the 7th ACM conference on Recommender systems*, pp. 17–24, New York, New York, USA, 2013. ACM Press.
- [5] Linas Baltrunas, Marius Kaminskis, Bernd Ludwig, Omar Moling, Francesco Ricci, and Aykan Aydin. InCarMusic : Context-Aware Music Recommendations in a Car. In *[EC-Web2011]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 *RecSys Decision 2013: Proceedings of the RecSys 2013 Workshop on Human Decision Making in Recommender Systems*, pp. 21–28, 2013.
- [7] Songhua Xu, Hao Jiang, and Francis C.M. Lau. Personalized online document, image and video recommendation via commodity eye-tracking. In *[RecSys2008]Proceedings of the 2008 ACM conference on Recommender systems*, p. 83, New York, New York, USA, 2008. ACM Press.
- [8] Aleksandar Stupar and Sebastian Michel. PICASSO To Sing you must Close Your Eyes and Draw. In *[SIGIR2011]Proceedings of the 34th international ACM SIGIR conference on Research and development in Information*, pp. 715–724, 2011.
- [9] 奥健太, 山西良典. 土地被覆図からの景観要素抽出に基づく道路リンクの景観ベクトル化. 情報処理学会研究報告, 第 2017-DBS-1 巻, pp. 1–6, 2017.