

## 時間区間アノテーションの集約に基づく楽曲推薦システム

## A Music Recommendation System Based on Summarization of Temporal Annotations

真野 洋平<sup>†</sup>

Yohei Mano

高田 敏弘<sup>‡</sup>

Toshihiro Takada

## 1 はじめに

インターネットの普及とその商用利用の拡大に伴い、インターネットを介して容易に大規模楽曲データベースにアクセスできるようになり、ユーザが多くの音楽をいつでもどこでも手に入れられるようになった。その一方で、膨大で多様な楽曲データの中から好みの楽曲を手に入れることは容易ではない。標準的な楽曲検索システムではユーザが検索キーワードを入力しなければならず、どのようなキーワードで検索すれば好みの曲が発見できるのかわからず途方に暮れることもある。このような問題に対処するためにユーザの音楽的嗜好をシステムが自動的に推測し、その嗜好にあった楽曲を推薦する技術の研究が進められている [1, 2, 3]。

また、インターネットを介した音楽配信が普及するに伴い、ウェブ上で音楽について会話をする機会が増えた。例として、音楽や動画の視聴サービスはもちろんのこと、音楽ストリーミング（聴き放題）サービスやDTM（Desk Top Music：コンピュータを活用した音楽制作）のコミュニティなどの場面が挙げられる。これらのウェブサービスで会話が盛んに行われるようになったのには、単に音楽について感想を述べたいという思いだけでなく、音楽の感想や視聴経験を共有したいという思いを多くの人々が抱いているからだと考えられる。実際、そのような視聴経験の共有機能を持った音楽配信サービス（SpotifyやK.K.BOXなど）は急激にユーザ数を伸ばしており [4]、ウェブ上で音楽について会話をするユーザは今後益々増えていくと推測される。

音楽を介したウェブコミュニケーションで最もよく使用されているのは楽曲に付与されるコメントであろう。しかし、ウェブ上で音楽に対するコメントは相手に内容を伝えるににくいという課題がある。例えば、「1分47秒からのギターの音が変だ」「ラストのドラムのフィルインがカッコいい」といったような既存の時間指定のコメントを理解するには、その部分を探して聞く必要があり、そ

ういったコメントが複数あった場合には確認するのが面倒になってしまう。また、「ここがいいね」といった指示語も音楽に対しては使いづらい。作曲コミュニティでも音楽内容にコメントで意見を述べる際に細かい時間の指定が煩雑であったり、どの部分のどのような音楽内容（メロディ、構成、進行など）について意見をしているかが伝わりにくく議論が進まないといった課題がある。より円滑な音楽コミュニケーションを可能にするためには、音楽との関連付けや共有性、一覧性のある新たなコミュニケーション手法が必要であると考えられる。

そこで我々はウェブ上で音楽について円滑な会話、かつ視聴体験の共有ができたなら便利であろうというモチベーションの下、音楽に対して時間区間的にコメントや評価を付与できる時間区間アノテーションシステムを開発した。さらに、このシステムにより複数のユーザが付与したアノテーション（コメントや評価）を取りまとめれば、その楽曲内で注目されている部分が変わるのではないかと考え、アノテーション領域の集約による楽曲のダイジェスト生成のアルゴリズムを考案した。本稿では、この方法で生成されたダイジェストはサビ抽出や構造抽出といった音楽音響信号の解析だけでは取り出すことのできない人の感性が反映されている点に着目し、不特定多数ユーザによるアノテーションの集約により生成されたダイジェストに基づいた楽曲推薦システムを提案する。人の感性が反映されたダイジェストとユーザの音楽的嗜好を直接比較することで、よりユーザの嗜好に適し、かつ思いがけない発見性のある楽曲推薦を目指す。

本稿は以下のように構成される。2章では我々が提案する時間区間アノテーションの集約に基づいた楽曲推薦手法について説明する。3章では音楽推薦の従来手法と提案手法を比較し、それぞれの長所短所について述べる。4章では提案手法が従来手法より優れていることを示すための評価実験について説明する。5章では本研究の進捗状況と今後取り組むべき課題について述べる。最後に6章でまとめとする。

<sup>†</sup>名古屋大学大学院 情報科学研究科<sup>‡</sup>NTT コミュニケーション科学基礎研究所

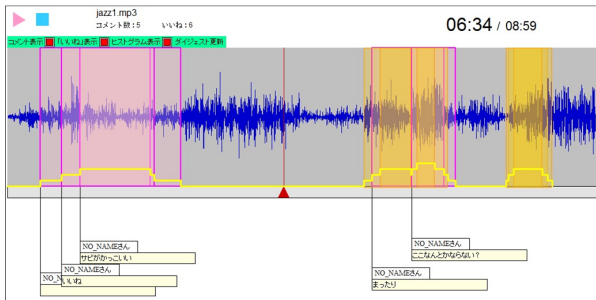


図 1: 時間区間アノテーションシステムのスクリーンショット

## 2 本研究のアプローチ

この章では、時間区間アノテーションシステムとダイジェストに基づいた楽曲推薦の概要を述べた後、ダイジェスト生成、類似度計算について述べる。

### 2.1 時間区間アノテーションシステム

本稿における時間区間アノテーションとは、音楽に対して時間区間的に付与されたコメントや評価 (Facebookなどで使用される「いいね」) などの情報を表す。本研究で開発した時間区間アノテーションシステムは、複数のユーザが音楽波形に時間区間アノテーションを自由に付与できるシステムである。アノテーションの付与方法は2種類実装しており、クリック&ドラッグで範囲を指定するものと音楽を聞きながら再生バーに合わせてクリックして範囲の始端と終端を決定するものがある。時間区間アノテーションの重なり具合をグラフ化し、盛り上がり度を表示するヒストグラムやコメント部分にマウスを乗せると対応する領域が強調される機能など、視覚的に理解しやすくなるような機能を実装している。システムのスクリーンショットを図1に示す。

### 2.2 ダイジェストに基づいた楽曲推薦

本研究では楽曲内の話題になっている部分をダイジェストと定義する。複数ユーザによるアノテーションが多く重なり合っている部分がダイジェストになる。このようなダイジェストを用いた本研究の提案手法の概要を図2に示す。楽曲推薦を希望するユーザ A に対して以下の手順で楽曲推薦を行う。

- (1) 複数ユーザのアノテーション領域を要約して、その楽曲のダイジェストを生成する。
- (2) 各楽曲のダイジェスト領域の音声信号とユーザ A が付けたアノテーション領域の音声信号を比較し、類

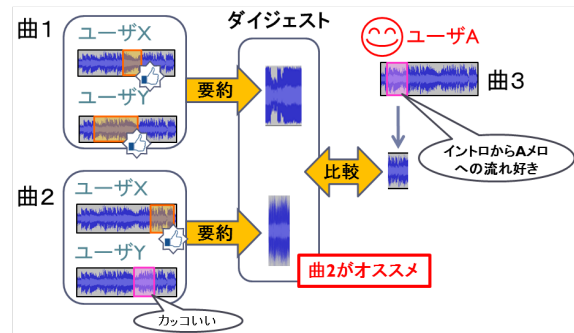


図 2: 時間区間アノテーションの集約に基づく楽曲推薦システムの概要

似度を計算する。アノテーション領域の音声信号が検索クエリになる。

- (3) 計算された類似度に合わせてダイジェストに順位を付け、順位上位のダイジェストとそのダイジェストの元である楽曲を推薦結果としてユーザ A に提示する。

#### 2.2.1 ダイジェスト生成

ダイジェストは先にも述べたように楽曲内の話題になっている部分を表し、複数のユーザのアノテーション領域の重なりから生成する。ダイジェスト生成の概要を図3に示す。

ある一曲のダイジェストの生成を以下の手順で行う。

- (1) 一曲内のアノテーションの重なり度合いを調べ、重複数ヒストグラムを生成する。各ビンの値がアノテーションの重複数を表している。
- (2) ヒストグラムのビンの値と個数から平均重複数を求める。

$$\text{平均重複数} = \frac{\text{全ビンの値の和}}{\text{値が1以上のビンの個数}} \quad (1)$$

- (3) ヒストグラムの中で平均重複数を超えるビンを抽出し、その中で隣接するビンをグループとしてまとめ、それらグループがダイジェストの候補となる。
- (4) 各グループの幅に比べて隣接するビンの幅が二分の一以下であれば、そのグループに統合する。グループ同士が接したら、グループを統合する。統合できるビンがなくなったら終了する。この時点で残っているグループがその楽曲のダイジェストになる。
- (5) 統合の結果生成されたダイジェストの始端と終端の周辺の音声信号を分析し、周波数スペクトルの変化

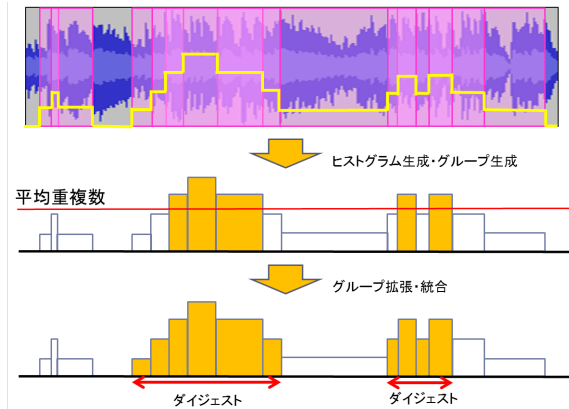


図3: ダイジェスト生成の概要

などを見て、音楽的に区切りの良いようにダイジェストの始端と終端の位置を調整する。

### 2.2.2 類似度計算

前節の手法で生成した各楽曲のダイジェスト領域の音声信号と個々のユーザのアノテーション領域（クエリ領域）の音声信号との類似度を [5] の手法に沿って計算する。この手法を用いることでクエリとの全体的な曲調が似ているダイジェストを精度よく見つけることができる。また、この手法は音声信号に多少の欠損があっても頑健な類似度計算が可能であり、楽曲から部分的に切り出されたクエリ領域やダイジェスト領域の音声信号に対しても頑健な類似度計算が期待できる。

各楽曲のダイジェスト領域とクエリ領域の音声信号から MFCC (Mel-Frequency Cepstral Coefficients) と呼ばれる周波数スペクトル特徴量を抽出し、得られた MFCC ベクトル集合を k-means 法でクラスタリングすることで、クエリ領域とダイジェスト領域の音声信号を特徴量分布の集合として表現する。クエリと各ダイジェストの特徴量分布間の距離を EMD (Earth Mover's Distance) を用いて計算する。EMD は線形計画問題の一つである輸送問題における輸送コストの最適解を分布間の距離により求める方法であり、類似画像検索の分野などに応用されている [6]。EMD により求めた分布間の距離をクエリ/ダイジェスト間の距離とする。EMD の値が小さい（クエリ/ダイジェスト間の距離が近い）順に順位付けを行い、ランキング上位の楽曲を推薦結果として提示する。

以上で述べた時間区間アノテーションシステムとダイジェスト生成、類似度計算の手法を組み合わせたものが時間区間アノテーションの集約に基づく楽曲推薦システムである。

	協調	内容ベース	提案手法
正解率	○	×	◎
セレンディピティ	○	×	◎
文化的背景、市場人気	○	×	○
cold-start問題	×	△	△
被覆率	×	○	○
利用者数	×	○	△

図4: 提案手法と従来手法との比較

## 3 従来手法との比較

本章では提案手法であるダイジェストに基づいた楽曲推薦と従来手法とを比較し、それぞれの長所短所について考察する。

楽曲推薦の従来手法として協調フィルタリング (collaborative filtering) と内容ベースフィルタリング (content-based filtering) の2つが挙げられる [7, 8]。協調フィルタリングはあるユーザに対して、好みの似た他のユーザが好んでいる楽曲を推薦する手法である [9]。例として「この曲を聴いているユーザはこんな曲も聴いています」といった形式で推薦され、この手法は Amazon.com や Last.fm などの商用サービスでよく使われている。内容ベースフィルタリングはユーザが好む楽曲と音楽内容が類似している楽曲を推薦する手法である [10]。利用される音楽内容は音楽音響信号に基づくものやメタデータに基づくものがある。音楽配信サービス Pandora では専門家が音楽内容を約400もの観点から分析し、点数付けを行い推薦に活用するという仕組みをとっており、高い精度で推薦ができるが、コストがかかるという課題がある。

これらの従来手法の長所と短所を [7, 8] に基づいて音楽推薦で重要な6つの項目として図4に示すようにまとめ、提案手法との比較検討を行った。

### 3.1 正解率

正解率はユーザの関心への適合が推薦結果とテスト用データで一致した割合を表す。ユーザの嗜好にあった結果をどの程度推薦できているかとも言える。

協調フィルタリングでは他のユーザの評価を基に推薦を行うため、比較的正確率の高い推薦が期待できる。内容ベースフィルタリングでは、様々な音楽内容に基づいて推薦を行うが、必ずしもユーザが期待している音楽内容を扱っているとは限らず、安定して正確率の高い推薦ができない。

提案手法では他のユーザの評価の集約結果であるダイ

ジェストを利用して他のユーザの評価を取り入れている。また、ユーザがその時点で好きだと感じた音楽の一部をクエリとして直接推薦に活用しており、ユーザの求めている音楽内容を取り入れやすい。これらの点から提案手法は従来手法に比べて、より正解率の高い推薦が行うことができるのではないかと考えられる。

### 3.2 セレンディピティ

セレンディピティとは元々は「偶然によって思いがけず価値あるものを発見する能力」を意味する [11, 12]。楽曲推薦においていえば、ユーザが想定していなかった新たなジャンルやアーティストなどといった新規性があり、かつユーザが興味を持ってくれそうな楽曲を推薦することができる能力を表す。

協調フィルタリングはユーザ自身が知らないジャンルやアーティストでも他のユーザの知識を通じて知ることができるためセレンディピティの観点では有利といえる。内容ベースフィルタリングはユーザ自身が知っている音楽の内容や知識に沿った楽曲に推薦対象が限定されてしまうことが多い。

提案手法ではダイジェストという形で他のユーザの知識や興味を活用している。また、ダイジェストは類似度計算の結果として推薦されるが、それに伴ってユーザが知ることができるダイジェストの元の曲に関してはシステムが関知することなく提示されるため、提案手法では十分な新規性や意外性を与えることができると考えられる。例として、楽曲中の Jazz 的なフレーズの部分がダイジェストになっているが曲自体は Rock のジャンルに含まれる楽曲 R があるとす。Jazz を好んで聴くユーザが Jazz の曲にアノテーションを付けた際に、偶然 R のダイジェストが推薦されたときにユーザは R のダイジェスト以外の部分を聴く機会が与えられる。R のダイジェストは Jazz というユーザの興味に適しており、このダイジェストをきっかけに Rock というジャンルに興味を持つ可能性がある。このようにダイジェストを活用することで、本来はトレードオフの関係にある正解率とセレンディピティを両立させた推薦が可能になると考えている。

### 3.3 文化的背景、市場人気

ユーザの音楽的嗜好はユーザの文化的背景とその時々市場人気に反映されることが多いため、楽曲推薦に文化的背景や市場人気を考慮できる要素は重要である。

協調フィルタリングは他のユーザの評価を基に推薦を行うため、文化的背景や市場人気を考慮した推薦が可能

である。内容ベースフィルタリングはユーザ自身のプロフィールとデータベースの音楽内容を比較するだけであるから、文化的背景や市場人気はほとんど反映されない。

提案手法では他のユーザの評価を基に生成されたダイジェストを利用するため、文化的背景や市場人気を考慮した推薦が可能である。

### 3.4 cold-start 問題

cold-start 問題 (start-up 問題や ramp-up 問題とも呼ばれる) は 2 種類に分けられる [7]。一つは新たにシステムを利用し始めたユーザに適切に推薦を行う難しさであり、もう一つは推薦対象として新たにシステムに導入された楽曲を推薦する難しさである。

前者については協調フィルタリングでは、新規ユーザの情報が少なく、他のユーザのとの類似性を判定できないため、適切な推薦を行うことが難しい。内容ベースフィルタリングでは、自身のプロフィールを入力するタイプのものであれば比較的適切に推薦ができる。提案手法では、ダイジェストが準備できていれば、新規ユーザでも何かしらの楽曲の一つでもアノテーションを付与するだけで特に問題なく推薦を行うことができる。

一方、後者の新規楽曲に対する問題については、協調フィルタリングではユーザの評価を利用するため、まだ誰も評価していない楽曲は推薦することができない。内容ベースフィルタリングではユーザ自身のプロフィールがあれば、新規の楽曲に対しても適切に推薦ができる。提案手法では新規楽曲にはダイジェストを生成できないため、推薦対象となりえない。しかし、楽曲登録時に仮のダイジェスト (イントロやサビなど) を用意することで推薦対象に加えることはでき、この問題に対処することが可能となる。

### 3.5 被覆率

全楽曲のうち、評価値の予測が可能な楽曲の割合を表す。

協調フィルタリングでは、まだ誰も評価していない楽曲は推薦の対象にできないため、被覆率を 100% にするのは困難である。内容ベースフィルタリングでは楽曲の特徴を手掛かりにすべての楽曲に対して推薦すべきかどうかを判定できるため、被覆率は 100% になる。

提案手法では、ダイジェストが生成されていない楽曲は推薦の対象にできない。しかし、cold-start 問題でも述べた仮のダイジェストを用意することで、未評価楽曲であってもダイジェストが存在し推薦対象に含めることができるため、被覆率を 100% にすることができる。

### 3.6 利用者数

利用者が少ない際に適切に推薦ができるかどうかという観点である。

協調フィルタリングでは、他のユーザの評価を参照するので利用者数が少ないと機能しない。また、利用者が他にいない状況では、適切な推薦もできないため、新たな利用者を増やすことも難しいという問題がある。内容ベースフィルタリングでは、ユーザ自身のプロフィールとデータベースとの比較であるため、たとえ利用者が一人であっても推薦は可能である。

提案手法では、利用者数が少ないと他のユーザの評価からダイジェストを生成できない問題があるが、仮のダイジェストを用意することで利用者が一人でもシステムとしては機能する。利用者数が少ない場合にユーザを納得させる推薦を行うためには、仮のダイジェストの質を高める必要があり、サビや楽曲構造の抽出技術を取り入れることになる。

以上の考察から「音楽推薦で重要とされる6種類の要素において提案手法が従来手法よりも優れている、あるいは同等の性能を有する」という仮説を立てることができる。

## 4 仮説の実験的検証

前章で述べた6種類の要素の内、正解率とセレンディピティについて実験的検証を行い仮説が正しいことを証明する。提案手法と従来手法(強調フィルタリング、内容ベースフィルタリング)による楽曲推薦の正解率とセレンディピティ度合の測定を目的として実験を実施する予定である。実験の手続きの概要を図5に示す。

### 4.1 実験方法

実験に用いる楽曲は1990～2000年の日本のヒット曲50曲である。この50曲をデータベース用とテスト用にランダムに25曲ずつ分けておく。以下に被験者実験の基本的な流れを示す。

- (1) 被験者はデータベース用楽曲を視聴してもらい、その中から好みの曲を数曲選んでもらう。その際に好みの曲が既に知っている曲だったのか、初めて知った曲なのかについても区別する。ここで選ばれた楽曲は正解楽曲データとなる。
- (2) 被験者にテスト用楽曲を視聴してもらい。視聴した楽曲の中で、気に入った部分があれば時間区間アノテーションシステムを利用してアノテーションを付

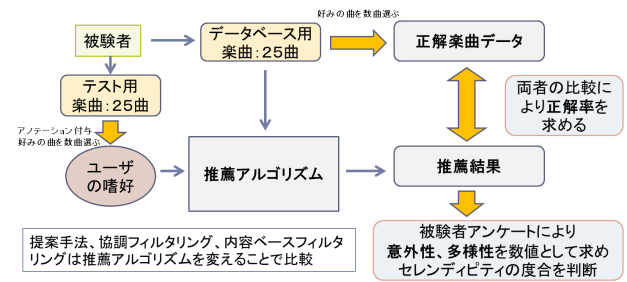


図5: 実験手続きの概要

与してもらい。また、テスト用楽曲の中からも好みの曲を数曲選んでもらう。

- (3) (2) から得られた被験者の嗜好データを基に推薦アルゴリズムに従ってデータベース用楽曲の中から楽曲を推薦し、その結果に対して被験者アンケートを実施する。

正解率は正解楽曲データと推薦結果との比較により求める。被験者アンケートでは推薦結果に対して意外性、多様性などを数値として求め、セレンディピティの度合を判断する。提案手法と協調フィルタリング、内容ベースフィルタリングの比較は(3)の推薦アルゴリズムを変えることで比較を行う。

### 4.2 実験の準備

本実験を行う前に準備しておかなければならないデータが2種類ある。一つはダイジェスト生成の為のアノテーションデータで、もう一つは協調フィルタリングの為の評価データである。これらのデータがなければ、前節の(3)の楽曲推薦で参照するダイジェストや他のユーザの評価が存在しないため、推薦を行うことができない。

そのため10人の大学生に協力してもらい、データ収集を行った。実験に使用する50曲を視聴してもらい、その中から好みの曲を10曲選ぶ。これは協調フィルタリングのための評価データとして利用する。さらに選んだ10曲に対してどこが気に入った部分なのかを時間区間アノテーションシステムによってアノテーションを付与してもらい、残りの40曲に対しても部分的に気に入った箇所があればアノテーションを付けてもらう。こちらは提案手法のダイジェスト生成の為のアノテーションデータとして使われる。

## 5 考察

この章では、本研究の進捗状況と今後取り組むべき課題について述べる。

現在、2章で述べた時間区間アノテーションシステムとダイジェストに基づいた楽曲推薦手法を組み合わせた楽曲推薦システムを開発している。将来的には、このシステムを不特定多数のユーザが利用できるように Web サービスとして公開する予定である。

時間区間アノテーションシステムはコンピュータ上で作曲した音楽に対してウェブ上で意見交換を円滑に行うために開発したものであり、音楽的知識のあるユーザが楽曲の内容について議論するという利用方法を想定していた。「ここからここまでは直したほうがいい」といった指示語を使ったコメントが使える点やコメントの重なり具合からどの部分で議論が盛んであるかが一目でわかるという点がシステムの特徴である。これらの特徴により音楽批評の際の時間指定の煩雑さやどの部分のどのような音楽内容に対して意見を述べているのかが伝わりにくいなどの課題を解決できている。不特定多数のユーザにもこのシステムを利用してもらいたいと考えており、音楽に時間区間アノテーションを付与することへの動機付けをどのように行うかが今後の課題となっている。

ダイジェストに基づいた楽曲推薦システムはダイジェスト生成と類似度計算の要素技術はそれぞれ完成しており、これらを組み合わせる段階にある。ダイジェスト生成については、アノテーション重複に合わせてダイジェストを作ることはできているが、音楽的に区切りの良いように音楽解析技術を取り入れてダイジェストの始端と終端を調整する手法を検討していく必要がある。類似度計算については、MFCC を特徴量として用いたことにより、推薦結果は声質の似ている歌手の曲や同じような楽器で演奏している曲など、周波数成分が似ている楽曲が上位に来る傾向が見られた。現時点では特徴量は MFCC のみを使用しているが、MFCC だけでは周波数的観点から見た推薦結果を提示することしかできない。より柔軟な楽曲推薦を行うために、楽曲構造やテンポの抽出技術などを取り入れ、類似度計算に反映させていく予定である。

4章で述べた被験者実験に関しては、提案システムを完成させ、比較対象となる協調フィルタリングと内容ベースフィルタリングの推薦アルゴリズムを準備した後、実施する予定である。

## 6 おわりに

本稿では、時間区間アノテーションの集約に基づく楽曲推薦システムを提案した。提案手法は複数のユーザによって音楽に付与された時間区間アノテーションの集約することで音楽音響信号の解析では取り出すことのでき

ない部分も含めた楽曲のダイジェストを生成できるというアイデアに基づく。提案手法と音楽推薦の従来手法との長所短所の比較検討により、提案手法が正解率、セレンディピティなどの6つの観点で優れている、または同等の性能を有することの仮説を立てた。

今後は、4章で述べた被験者実験を行い仮説の真偽を検証する予定である。

## 参考文献

- [1] Typke, R., Wiering, F., & Veltkamp, R. C. "A survey of music information retrieval systems," Proc. of the ISMIR, pp. 153-160 (2005).
- [2] Celma, O. "Music recommendation and discovery in the long tail," Ph.D. dissertation, Univ. Pompeu Fabra, Barcelona, Spain (2008).
- [3] 吉井和佳, 後藤真孝. "音楽推薦システム," 情報処理(情報処理学会誌), Vol. 50, No. 8, pp. 751-755 (2009).
- [4] 柴 那典. "「聞き放題」だけでは音楽ストーリーミングサービスが成功しない理由," <http://www.drillspin.com/articles/view/571> (2013年6月現在).
- [5] Logan, B., & Salomon, A. "A music similarity function based on signal analysis," In IEEE International Conference on Multimedia and Expo, pp. 745-748 (2001).
- [6] Rubner, Y., Tomasi, C., & Guibas, L. J. "The earth mover's distance as a metric for image retrieval," International Journal of Computer Vision, Vol. 40, No. 2, pp. 99-121 (2000).
- [7] 神島敏弘. "推薦システムのアルゴリズム (1)-(3)," 人工知能学会誌, Vol. 22, No. 6 - Vol. 23, No. 2 (2007-2008).
- [8] Yoshii, K., Goto, M., Komatani, K., Ogata, T., & Okuno, H. G. "An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model," Audio, Speech, and Language Processing, IEEE Transactions, Vol. 16, No. 2, pp. 435-447 (2008).
- [9] Su, X., & Khoshgoftaar, T. M. "A survey of collaborative filtering techniques," Advances in Artificial Intelligence (2009).
- [10] Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M. "Content-based music information retrieval: Current directions and future challenges," Proc. of the IEEE, Vol. 96, No. 4, pp. 668-696 (2008).
- [11] 奥健太, 服部文夫. "セレンディピティ指向情報推薦のためのフュージョンベース推薦システム," 知能と情報(日本知能情報ファジィ学会誌) - 特集: Web インテリジェンスとインタラクション II -, Vol. 25, No. 1, pp. 524-539 (2013).
- [12] 澤泉重一, 片井修. "セレンディピティの探求: その活用と重層性思考," 角川学芸出版 (2008).