

時間区間アノテーションの集約に基づくダイジェストを対象とする楽曲推薦システム

真野 洋平¹ 高田 敏弘^{2,1} 齋藤 洋典¹

概要：本論文は、複数のユーザによって楽曲に付与された時間区間アノテーション（コメントや評価）の集約によって得られる楽曲のダイジェストを用いた、楽曲推薦システムを提案する。本システムは、ユーザが楽曲内の気に入ったフレーズを検索クエリとして選択すると、そのクエリと各楽曲のダイジェストとの類似度を計算し、クエリによく似たダイジェストを持つ楽曲を推薦する。本手法は、時間区間アノテーションに基づいたダイジェストに着目することで、部分的フレーズに対する楽曲推薦を可能にする。このダイジェストは、計算機による構造抽出とは異なり人の音楽に対する感性を反映しているため、よりユーザの嗜好に適した推薦を行うことができると考えられる。楽曲推薦実験を通じて、提案手法が正解率とセレンディピティ（思いがけず価値のあるものを発見する能力）に関して優れた推薦能力を持つことを示す。

1. はじめに

インターネット技術の発展とその商用利用の拡大に伴って、ユーザは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、大量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。しかし、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効率的に見つけるのは非常に大変である。標準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楽曲を見つけ出すことができるのか分からず途方に暮れてしまうこともある。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、その嗜好に適した楽曲を推薦する楽曲推薦技術が盛んに研究されている。

楽曲推薦の従来手法は、協調フィルタリングと内容ベースフィルタリングの2つに大別でき、両手法は相補的な性質を持っている。協調フィルタリングは、基本的にはあるユーザに楽曲を推薦する際に、他のユーザの評価を利用して楽曲を推薦する手法である[1]。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、Amazon.com や iTunes Music Store などの商用サービスでもよく利用されている。内容ベースフィルタリングはユーザが好む音楽内容が類似している楽曲を推薦する手法である。音楽内容としては主に音楽音響信号を解析

して抽出される音響特徴量や楽曲のメタデータが用いられる[2]。さらに近年では、両手法の欠点を解消するため、両手法の特徴を兼ね備えたハイブリッド型フィルタリングによる楽曲推薦に注目が集まっている[3]。

これまでの楽曲推薦では楽曲単位の検索クエリを基にして、楽曲単位の推薦結果を提示する形式が一般的であり、楽曲の部分的フレーズを対象とした推薦はほとんど行われてこなかった。音楽を聴いていて、気に入ったフレーズに出会ったときに、そのフレーズの音楽内容と類似したフレーズを持つ他の楽曲を推薦することができたら、それは多くの音楽ユーザにとって有用であると考えられる。しかし、ユーザが求めるフレーズを無数の楽曲の中から見つけ出すことは難しい問題である。楽曲を1曲毎に細かく分析することによって、類似フレーズを見つけることは可能ではあるかもしれないが現実的ではない。また、見つけた中から更に適切なフレーズを選んで推薦しなければ、結果に対してユーザを納得させることはできない。このような推薦対象の生成と推薦能力の課題があったために、これまで部分的なフレーズに対する楽曲推薦はほとんど行われてこなかった。

そこで、本研究では複数のユーザの時間区間的なコメントや評価を基にして、各楽曲のダイジェストを抽出することで多くの人々が気に入っているフレーズの見つけ出すことが可能になることに注目した。このダイジェストを用いることで、推薦対象の生成課題を解消し、楽曲の部分的なフレーズに対する楽曲推薦を実現した。また、このダイジェストは人の音楽に対する感性を反映しているため、従

¹ 名古屋大学大学院 情報科学研究科
Graduate School of Information Science, Nagoya University
² NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories

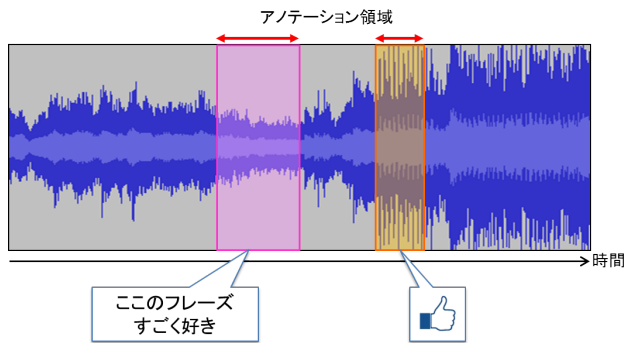


図 1 時間区間アノテーション

来の音楽推薦手法に比べて、よりユーザの嗜好に適した楽曲推薦が可能になると考えられる。

提案手法は、ダイジェスト抽出に協調フィルタリング、類似度計算に内容ベースフィルタリングの特徴を取り込んでおり、一種のハイブリッド型フィルタリングとしてとらえることができる。そのため協調フィルタリングと内容ベースフィルタリングの2つの従来手法に比べて、提案手法は優れた推薦結果を提示することができると考えられる。本論文では正解率とセレンディビティ（思いがけず価値のあるものを発見する能力: Serendipity）の項目に関して提案手法と2つの従来手法との比較検討を行い、この2項目に関して提案手法はより優れた推薦能力を持つことを実験的に検証する。

2. 本研究のアプローチ

本章では、提案手法である時間区間アノテーションの集約に基づくダイジェストを対象とする楽曲推薦について説明する。この手法によって、ある楽曲内の部分的なフレーズが気に入った場合に、そのフレーズに類似したフレーズを持つ他の楽曲を推薦することが可能となる。時間区間アノテーションの集約によるダイジェストは計算機による構造抽出とは異なり、人の音楽に対する感性を反映していることから、提案手法による推薦結果は従来手法に比べてより適切な結果を提示することが期待できる。

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 (Annotation) とは、音楽や映像のように時間的な流れを持つようなコンテンツに対して時間区間的に付与したコメントや評価 (Facebook などで使用される「いいね」) などの情報として定義する (図 1)。

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、コミュニケーションを図ることが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

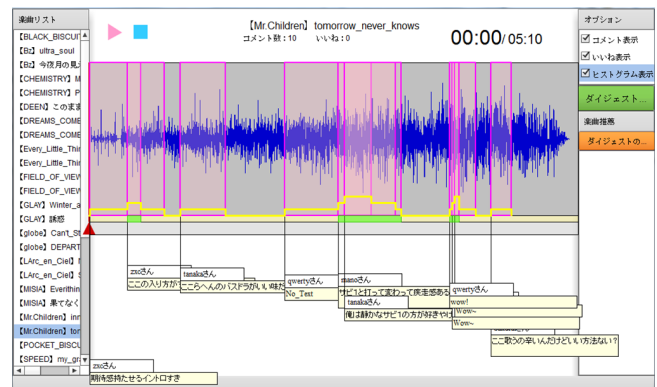


図 2 システムのスクリーンショット

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。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2に示す。

2.2 ダイジェストに基づいた楽曲推薦

時間区間アノテーションを一つの楽曲に複数のユーザが付与していくと、アノテーション区間が重なった部分が出てくる。アノテーションが多く重なっている区間は多くのユーザが注目しており、話題になっている箇所であるとみなすことができる。このアノテーションの集約によって得られる楽曲の一部分を本研究ではその楽曲のダイジェストとして扱い、楽曲推薦の際の検索対象として用いる。

提案手法の概要を図 3 に示す。時間区間アノテーションシステムを利用して楽曲にアノテーションを付与したユーザ A に対して、以下の手順で楽曲推薦を行う。

- (1) ユーザ A が付与したアノテーション領域の音響信号を検索クエリ (Query) として切り出す。
- (2) 検索クエリの音響信号と各楽曲のダイジェスト領域の音響信号を比較し、それぞれの類似度を計算する。
- (3) 計算された類似度により各ダイジェストに順位を付け、順位が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。

各楽曲のダイジェストは他のユーザが既に付与したアノテーションを集約して逐次的に抽出され、ダイジェストデータベースが更新される。推薦結果からはユーザはダイジェストを聴くことができ、気に入った場合にはその元楽曲を聴くことができるようになっている。これにより推薦結果の中から効率良く、ユーザ自身の好みに合った楽曲を見つけることができる。

ユーザはこの推薦結果から別の楽曲を聴き、アノテーションを付与することで、また新たな楽曲推薦を得る。この一連の推薦手順を複数のユーザが行っていくことで、ダ

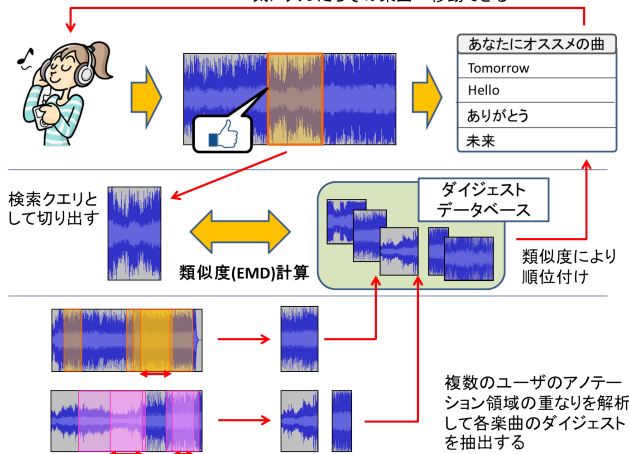


図3 提案手法の概要

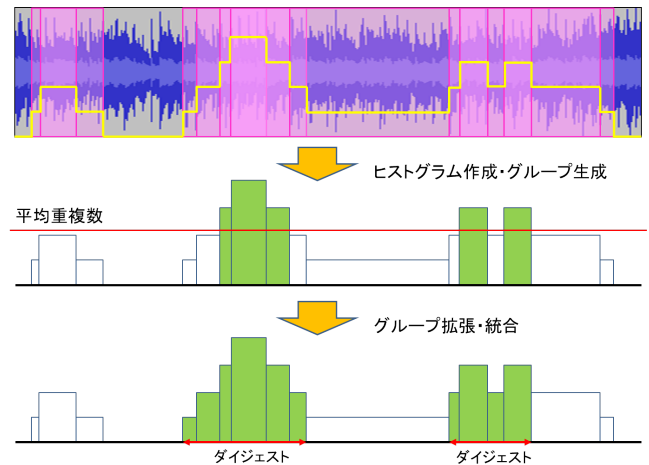


図4 ダイジェスト生成の概要

イジェストが洗練されていき、推薦結果もより納得のいくものになっていくというポジティブスパイラルを回すことが期待できるという点も提案手法の特徴の一つである。

2.2.1 ダイジェスト生成

ダイジェストは楽曲内の話題になっている部分を表し、複数のユーザーの時間区間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4に示す。

ある楽曲Xに対するダイジェストの生成を以下の手順で行う。

- (1) 楽曲X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。各ピンの値がアノテーションの重複数を表している。
- (2) ヒストグラムのピンの値と個数から平均重複数を求める。

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

- (3) ヒストグラムの中で平均重複数を超えるピンを抽出し、その中で隣接するピンをグループとしてまとめ、それらの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ピ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。統合できるピ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楽曲Xのダイジェストになる。

楽曲Xに新たなアノテーションが付与された場合、上記の手法で逐次的にダイジェストを更新していく。

2.2.2 類似度計算

クエリ領域の音響信号と各楽曲のダイジェスト領域の音響信号との類似度を[4]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音色的によく類似し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、この手法は楽曲をクラスタ集合として表現しているため、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

スト領域の音響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)[5]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によって分布間の距離を求める方法であり、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。

3. 従来手法との比較

本章では楽曲推薦の代表的な従来手法である協調フィルタリングと内容ベースフィルタリングの長所と短所について、楽曲推薦において重要な項目である「正解率」と「セレンディピティ」の観点で論じ[7], [8]、さらに提案手法との比較検討を行う。

3.1 正解率

正解率はユーザーの関心への適合が推薦結果とテスト用データとで一致した割合を表す。ユーザーの嗜好に合った楽曲をどの程度推薦できているとも言える。

協調フィルタリングでは他のユーザーの評価を参考にして推薦を行うため、ユーザーの音楽的嗜好を形成する要因である文化的背景や市場の人気度を反映することができ、比較的正解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、音響特徴量やメタデータなどの様々な音楽内容に基づいて推薦を行うが、必ずしもユーザーが期待している

音楽内容を扱って推薦するとは限らず、安定して正解率の高い推薦は難しい。さらに文化的背景や市場人気を取り入れることができないため、協調フィルタリングに比べると正解率は劣ってしまう。

提案手法では、他のユーザの評価を集約した結果であるダイジェストを参考にしており、他のユーザの評価を取り入れているため、文化的背景や市場人気を反映できる。また、ユーザがその時点で好きだと感じた音楽のフレーズを指定して検索クエリとして直接推薦に活用しているため、ユーザが楽曲推薦に求めている音楽的内容を取り入れやすい。従来手法では好きなフレーズとそうでないフレーズを持つ楽曲であっても1曲として扱っているため、ユーザのここが好きであるという強い嗜好を取り入れることはできていなかった。以上の特徴から、提案手法は従来手法に比べて、より正解率の高い推薦を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のあるものを発見する能力」を意味する [9], [10]。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や意外性があり、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のできる能力を表す。セレンディピティ要素を持つ推薦手法は、新たな楽曲との出会いを求めるユーザにとって良い推薦ができると考えられる。従来の楽曲推薦の研究では推薦精度の向上に重点が置かれており、セレンディピティに関してはあまり重視されてこなかった。

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内容ベースフィルタリングに比べて有利と言える。しかし、ある知名度の高いアーティストを好むユーザに対しては、同じアーティストの楽曲ばかり推薦されてしまうことが多くなる。内容ベースフィルタリングは多くの場合、ユーザが検索クエリを与えるため、ユーザ自身が知っている楽曲内容や知識に沿った楽曲やジャンル・アーティストに推薦対象が限定されてしまうことが多い。

提案手法はセレンディピティ要素を高める特徴を2つ持っている。一つはダイジェストという形で協調フィルタリングと同様に他のユーザの知識や興味を活用しているという点である。もう一つは、推薦結果としてダイジェストとその元になった楽曲がセットで提示されるという点である。ダイジェストはアノテーション集約による抽出や類似度計算などシステムによってクエリとの類似性を検証して選ばれたものであるが、ダイジェストの元楽曲全体（細かく言えばダイジェストで抽出された部分以外）についてはほとんど考慮されることなく選ばれて提示されるため、楽

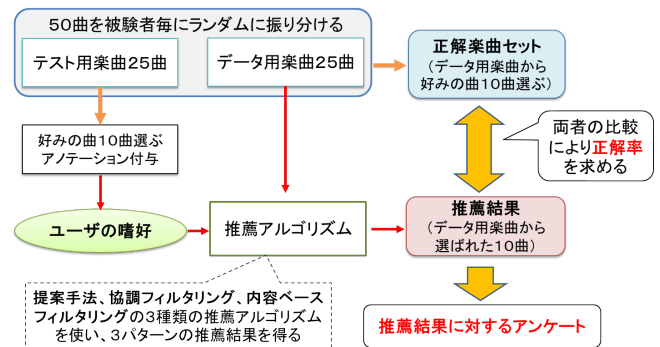


図5 実験1の手続き

曲によっては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として抽出されているが、曲全体としてはRockな雰囲気のある楽曲Rがあるとする。Jazzを好んで聴くユーザJがJazzの楽曲にアノテーションを与えた際に、偶然Rのダイジェストが推薦されるということは十分にありうる。その場合にユーザJにはRのダイジェスト以外の部分（Rock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Jの興味に適しており、本来Rockを聴かないユーザJに対して、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を与えることができる。楽曲単位で音楽を扱っている従来手法であれば、Jazzのフレーズが少しだけ含まれる楽曲RはRockというジャンルに正しく分類されるため、ユーザJに推薦されるようなことはない。

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共に向上させた推薦が可能になると考えられる。

4. 実験

本章では正解率とセレンディピティの観点で、提案手法の性能を確認するために実施した2つの実験について述べ、それぞれの結果に対する考察を行う。

4.1 実験1

実験1では提案手法による楽曲推薦は従来手法としてよく使われている協調フィルタリングと内容ベースフィルタリングに比べて正解率とセレンディピティの点で優れているという仮説を検証した。実験1の手続きの概要を図5に示す。実験には大学生20人に参加をしてもらった。実験に用いた楽曲は1990～2000年の日本のヒット曲50曲である。50曲もの音楽を聴くという作業を被験者が取り組みやすくするためにこのような選曲を行った。この50曲をデータ用楽曲とテスト用楽曲の2組にランダムに分けて実験1に用いる。以下に実験1の基本的な流れを述べる。

表 1 実験 1 の結果

項目	協調	内容ベース	提案手法
正解率	43.5 %	36 %	51 %
意外性	3.70	3.45	3.05
新規性	3.90	3.25	2.75
多様性	3.65	3.55	3.45
妥当性	3.70	3.35	4.05

- (1) 被験者はデータ用楽曲 25 曲を聴取し、その中から気に入った 10 曲を選ぶ。この 10 曲を正解楽曲リストとする。
- (2) 被験者はテスト用楽曲 25 曲を聴取し、その中から気に入った 10 曲を選ぶ。この各 10 曲の気に入っているフレーズに対して時間区間アノテーションを付与する。
- (3) (2) から得られた被験者の嗜好データを基に各推薦アルゴリズム（提案手法、協調フィルタリング、内容ベースフィルタリング）に従ってデータ用楽曲 25 曲の中から楽曲を 10 曲ずつ推薦し、3 パターンの推薦結果を得る。
- (4) 3 パターンの推薦結果を聴取した後、各推薦結果に対するアンケートを実施する。

正解率は正解楽曲リストの 10 曲と推薦結果の 10 曲を比較することで算出する。被験者アンケートでは各推薦結果に対して、意外性、新規性、多様性、妥当性に関する 1～5 の数値評価を実施した。

実験 1 の結果、得られた正解率とアンケートの各項目の評価値の平均を表 1 に示す。正解率と妥当性の項目で提案手法は高い値を示した。このことから提案手法は従来手法に比べてよりユーザの嗜好に合った楽曲を推薦できていると考えられる。多様性の項目には大きな差はなく、各手法とも同等の能力を有すると考えられる。一方、意外性、新規性の項目では従来手法は低い値を示した。これは、正解率が高いため、推薦結果の中に意外な楽曲や目新しい楽曲を推薦する余地が少なくなってしまうことが原因の一つと言える。また、評価の際にダイジェスト単位ではなく楽曲単位での聴取を行ったことも意外性、新規性の評価値が低くなった原因であると考えられる。

実験 1 の結果から正解率の観点では提案手法では提案手法は従来手法よりも優れているという仮説が正しいことを確認できた。その一方で、セレンディピティ性能を有するかどうかを実験 1 では十分に検証することはできなかった。そこで、実験 2 を行いダイジェストを聴かせることにより、提案手法は正解率とセレンディピティ性を両立させた推薦ができるのかを検証した。

4.2 実験 2

実験 2 ではダイジェストを聴かせる形式の結果提示を行い、推薦結果に対するアンケートから提案手法は正解率とセレンディピティを共に向上させることが可能かを検証

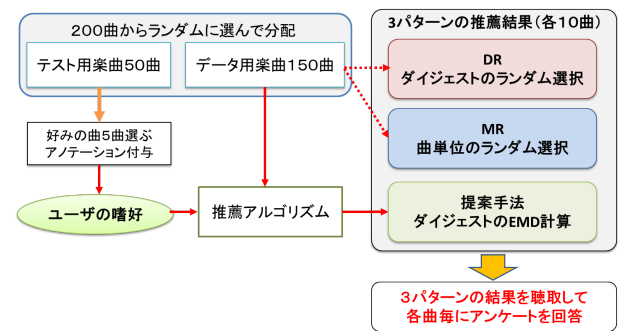


図 6 実験 2 の手続き

した。実験 2 には大学生 13 人に参加してもらった。実験に用いた楽曲は 1990～2000 年の日本のヒット曲 200 曲である。比較対象として意外性の高い推薦手法の一例としてランダムに曲を選ぶ方法を取り上げ、その推薦結果と提案手法による推薦結果とを比較することで検証を行う。この 200 曲をテスト用楽曲 50 曲とデータ用楽曲 150 曲の 2 組にランダムに分けて実験 2 に用いる。実験 2 の手続きを図 6 に示す。以下に実験 2 の手続きを述べる。

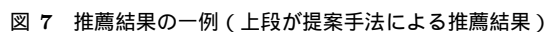
- (1) 被験者はテスト用楽曲 50 曲を聴取し、その中から気に入った 5 曲を選ぶ。この各 5 曲の気に入っているフレーズに対して時間区間アノテーションを付与する。
- (2) (1) から得られた被験者の嗜好データを基に提案手法の推薦アルゴリズムに従ってデータ用楽曲 150 曲の中から楽曲を 10 曲選ぶ。
- (3) 比較対象としてデータ楽曲 150 曲の中からランダムに 10 曲選んだ結果（MR と呼ぶ）と各データ楽曲のダイジェストの中から 10 曲選んだ結果（DR）を用意する。(2) で得られた提案手法の結果と合わせて計 3 パターンの推薦結果を被験者に提示する。
- (4) 3 パターンの推薦結果、計 30 曲（うち DR と提案手法 20 曲はダイジェスト）を被験者に聴取してもらい、被験者は各曲毎に対してアンケートの回答を行う。

被験者アンケートでは 4 つの質問に対して 1～10 の 10 段階評価を実施した。

実験 2 の結果、得られたアンケートの各項目の評価値の平均を表 2 に示す。最初に MR と DR を比較すると、Q1 の評価値は DR の方が大きい。同じようにランダムに選んで提示しているにも関わらず、ユーザに楽曲の頭から聴かせるよりもダイジェストとして馴染みのあるフレーズを聴かせた方がより知っている音楽だとユーザが気づきやすいため、ダイジェストを用いた楽曲推薦は提示した楽曲を既知だと感じる割合が高くなると考えられる。

次に DR と提案手法を比較すると、ユーザの嗜好に適した推薦ができているかを表す Q2 と Q3 の数値は提案手法が上回っていた。このことから提案手法はユーザの嗜好に

項目	DR	MR	提案手法
Q1：どの程度知っているか？	5.27	4.83	5.58
Q2：どの程度好きか？	6.01	5.95	6.28
Q3：普段聞く音楽とどの程度合っているか？	4.98	4.78	5.32
Q4：自力で出会うことができたと思うか？	5.43	5.28	5.41



実験2の結果から、提案手法はDRと比べて意外性の割合を大きく下げることなく、かつDRよりもユーザの嗜好に適した楽曲を推薦できているため、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性を共に向上させることを可能にする能力を持っていると言える。

本研究を通して、楽曲の部分的なフレーズに対する楽曲推薦システムを実現した。我々は部分的内容に対する推薦技術のネックとなっていた推薦対象の生成と推薦精度の向上を目的として、複数のユーザによる時間区間アノテーションを集約することで、その楽曲のダイジェストを抽出する手法に着目した。そして、ユーザが部分的に指定した検索クエリと各楽曲のダイジェスト間の類似度を EMD と呼ばれる距離尺度を用いて算出し、その数値に基づいて楽

今回提案した複数のユーザの意見を集約し、ダイジェストを抽出して推薦に活用するという推薦の枠組みは、音楽に限らず他のアイテム(本や動画など)にも適用可能であるかという点に関して検証することは非常に興味深い課題である。アイテムの内容の中で人が良いと感じる部分を取り出して推薦に活用するアプローチは、計算量の削減やユーザの興味に合わせやすい、更には新たなアイテムとの出会いをもたらすという観点からも推薦システムとして有用になるものと期待している。

- [1] X. Su and T. M. Khoshgoftaar, “A survey of collaborative filtering techniques”, *Advances in Artificial Intelligence* 2009, Vol. 4, pp. 1-20 (2009).
- [2] M. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney, “Content-based music information retrieval: Current directions and future challenges”, *Proceedings of the IEEE*, Vol. 96, No. 4, pp. 668-696 (2008).
- [3] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno, “An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model”, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 16, No. 2, pp. 435-447 (2008).
- [4] B. Logan and A. Salomon, “A music similarity function based on signal analysis”, In *IEEE International Conference on Multimedia and Expo*, pp. 745-748 (2001).
- [5] F. L. Hitchcock, “The Distribution of a Product from Several Sources to Numerous Localities”, *Journal of Mathematical Physics*, Vol. 20, pp.224-230 (1941).
- [6] Y. Rubner, C. Tomasi and L. J. Guibas, “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] 神鷹敏弘, “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] 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).
- [9] 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).
- [10] 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).