

# 事例ベース方式による内容に基づく音楽情報フィルタリング

学籍番号：90173068 西田研究室 竹川和毅

## 1 はじめに

内容に基づくフィルタリングを用いているシステムの多くは内容の形式がテキストである情報を対象としており、内容の形式が音楽そのものである情報を対象としたものは少ない。内容に基づくフィルタリングを行うにはユーザプロファイルが必要であるが、ユーザの嗜好からプロファイルを作成することはユーザにとって負担になり、音楽の内容をフィルタリングの対象とする場合にはその負担は特に大きい。あまり多くのユーザの評価は期待できない。本研究では、音楽データ中からフィルタリングにとって重要そうなデータを事例として獲得し、その事例を優先的にユーザに評価してもらうことで少ない評価でもある程度の性能の推薦を可能とする音楽情報フィルタリングシステムを提案する。

本研究では、MIDI データを用いて、事例ベース方式による音楽の内容に基づく情報フィルタリングシステムを構築する。フィルタリングの概要を図1に示す。まず、システムは音楽データの集合から重要そうなデータを事例として獲得し、ユーザはその事例に対して評価付けを行う。次に、評価付けされた音楽データの特徴量と評価を用いてユーザの嗜好を表すユーザプロファイルを作成する。そして、システムは推薦対象の音楽データの特徴量とユーザプロファイルと比較し、ユーザが気に入りそうな音楽データを推薦する。

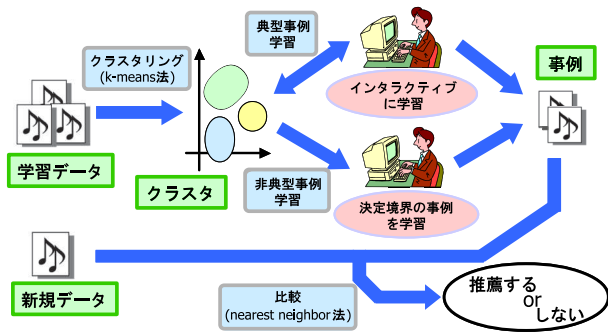


図1: 本研究のフィルタリングの概要

## 2 本研究のアプローチ

本研究では、以下の3つのアプローチにより事例学習を行う。

- “典型事例”, “非典型事例”の順の段階的な学習データの取得
- “典型事例”の学習における最低限の代表的な事例の学習
- “非典型事例”の学習における“典型事例”を補充するあいまいな事例の優先的な学習

学習を段階的に行うことで、ユーザの評価があまり得られなかった時は“典型事例”のみを学習してユーザプロファイルを作成し、ある程度の数の評価を得られた時は“非典型事例”も学習して、より正確なユーザプロファイルを作成する。そうすることで、ユーザの評価数が少ない場合でも、ある程度の推薦を可能にすることができる。

## 3 提案手法

### フィルタリングに用いる特徴量

本研究では音楽の特徴量として曲全体とCHごと(メロディ, コード, ベース及びドラムの4つのCH)の特徴量を用いる。表1にフィルタリングに用いる特徴量を示す。

表1: フィルタリングに用いる特徴量

曲全体	拍子, 調性, 平均テンポ, リズム, キー メジャーコードの割合, マイナーコードの割合 sus4 コードの割合
CH ごと	音色(メロディ) 平均音高差(メロディ, ベース, コード) 平均音長(メロディ, ベース, コード, ドラム) 平均音長差(メロディ, ベース, コード, ドラム)

### フィルタリングの仕組み

本研究では、ユーザプロファイルはユーザの曲に対する評価、コンテンツモデルは曲の特徴量となる。音楽データから学習した事例を用いてユーザプロファイルを作成し、推薦対象の音楽データと比較することによってフィルタリングを行う。事例の学習として、典型事例と非典型事例の2つの事例を学習する。

- 典型事例  
典型事例は、各クラスタの重心に、そのクラスタを代表するクラスを付けて事例として獲得したものであり、音楽データ集合のデータ分布を端的に表したような事例のことである。元の音楽データ集合から典型事例を獲得することで、元の音楽データ集合のデータ数よりも少ないデータ数で音楽データ集合を表すことができる。
- 非典型事例  
非典型事例は、音楽データ集合中のクラスタ間の決定境界付近にあるデータや、クラスタの重心から離れた例外を表しているようなデータを事例として獲得したものである。クラスタ間の決定境界付近はクラスタ分類の影響が少ないと考えられるため、非典型事例を用いて個別に判断する必要がある。また、例外も同様に非典型事例により個別に判断する。

## 4 システムの実装

上述のフィルタリング方式を持つプロトタイプシステムを Java と Java サブレットで実装した。

## 5 まとめ

情報フィルタリングの一般的なフィルタリング方式について検討し、本研究で用いる方式を決定した。そして、事例ベース方式による内容に基づく音楽情報フィルタリングの方式を提案した。提案方式により、学習データが少なくてもある程度の性能の推薦が可能となり、ユーザの負担を減らすことができると考えられる。

今後の課題として、提案システムの有効性を検証するための評価実験を行う。そして、実験結果を元に提案手法の修正・再検討をしていく。