

## ユーザ嗜好に基づく音楽情報検索システムにおける学習データ抽出手法

石先 広海	帆足 啓一郎	菅谷 史昭	甲藤 二郎
早稲田大学理工学研究科	KDDI 研究所		早稲田大学理工学研究科
〒169-8555 東京都新宿区	〒356-8502 埼玉県ふじみ野市		〒169-8555 東京都新宿区
大久保 3-4-1	大原 2-1-15		大久保 3-4-1
E-mail:ishizaki@katto.comm	Email:{hoashi,fsugaya}		E-mail:katto@waseda.jp
.waseda.ac.jp	@kddilabs.jp		

### 概要

ユーザ嗜好に基づく音楽情報検索システムでは学習データとしてジャンル情報を用いて検索対象楽曲の特徴空間を形成しているが、ジャンル情報では特徴空間内における楽曲分布に偏りが生じてしまう。そこで本稿では検索対象データに対して K-means 法に基づくクラスタリングを行うことで学習データの選別を行い、特徴空間を再構築することでこの問題を解消する。また K-means 法では初期値選定問題があり、その問題が検索精度に及ぼす影響を他のクラスタリング手法による結果との比較から、システムに最適なクラスタリング手法の考察を行った。

### Analysis of training data extraction for user preference based music information retrieval

<b>Hiroimi ISHIZAKI</b>	<b>Keiichiro HOASHI Fumiaki SUGAYA</b>	<b>Jiro KATTO</b>
Dept of Science and Engineering, Waseda University	KDDI R&D Laboratories, Inc.	Dept of Science and Engineering, Waseda University
3-4-1 Okubo Shinjuku-ku Tokyo 169-8555, JAPAN	2-1-15 Ohara Fujimino-shi, Saitama 356-8502, JAPAN	3-4-1 Okubo Shinjuku-ku Tokyo 169-8555, JAPAN

### Abstract

This paper proposes a training data extraction (TDE) method for user preference based music information retrieval (MIR). In conventional MIR, the biased distribution of songs in the feature space is a cause of MIR accuracy decrease. In this research, we propose a TDE method based on K-means clustering results. Experiment results prove that the proposed method solves the biased distribution problem, and improves MIR accuracy. However, there is a well-known prototype setting problem for K-means clustering. We have conducted further experiments to compare clustering algorithms, and examine the optimal clustering method for user preference based MIR.

## 1. はじめに

近年、音楽情報検索に関する様々な研究が行われている。しかし、既存の検索システムの多くは、検索対象となる楽曲データベースの中からある特定の楽曲を検索することを目的としており、ユーザが検索した音楽データについて何らかの事前知識(タイトル、演奏アーティスト、メロディ、等)があることが前提となっている。このようなシステムにおいて、大量の音楽データの中から未知の音楽情報を発見することは非常に困難である。そこで本稿では、ユーザの音楽的嗜好を学習し、ユーザが好むと思われる音楽情報を検索する手法に関して提案を行う。具体的には帆足らが提案したユーザ嗜好に基づく音楽情報検索システム[1]に関して、検索対象データの特徴空間の構築方法に関して提案、及び評価実験を行う。

## 2. ユーザ嗜好に基づく音楽情報検索システム

このシステムでは、まずツリーベクトル量子化手法(TreeQ)[2]に基づいて、学習データからベクトル量子化ツリーを形成する。次に形成されたツリーを用いて、すべての検索対象となる楽曲に対して、ベクトル量子化を行う。次に、ユーザに楽曲を提示し、試聴した楽曲に対して、「好き」、「嫌い」の評価を行ってもらい、その評価に基づいてユーザの嗜好ベクトル(以降ユーザプロファイル)を作成する。検索の方法はベクトル類似度に基づいており、ユーザプロファイルと検索対象楽曲の類似度を計算し、類似度の高いものから提示する。TreeQではツリー作成の際、正解カテゴリが付与されている学習データ(音楽データ)に基づいて、学習データが分類されるように最適化されたツリーを生成す

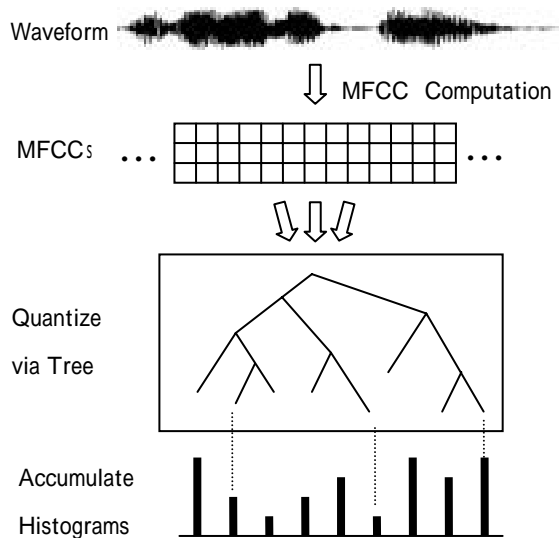


図1. ベクトル量子化手法概要

る。次に生成されたツリーに個々の音楽データ、あるいはカテゴリに属する全音楽データのMFCCのフレームを入力し、そのフレームが到達する葉(leaf)を求める。そして、各leafに到達したフレーム数を算出して得られるヒストグラムを各音楽データまたはカテゴリを表すベクトルとしている。図1にツリーベクトル量子化手法の概要を示す。

## 3. 既存システムの問題点

既存のシステムではTreeQの学習データとしてジャンル情報を用いている。これは、[1]でも述べられているが、すべてのユーザに対して普遍的な学習データを提供することができ、学習データの音響的な特徴も信頼できるという理論に基づいている。しかし、どれほどジャンル情報が正確で、TreeQに適していても、必ずしもすべての検索対象データベースに対して最適であることにはつながらない。現実的にはジャンル情報に基づいて構築された特徴空間を利用して検索対象楽曲をベクトル化すると特徴空間内においてある領域にベクトルが密集してしまう可能性がある。例えば、ある検索対象楽曲のほ

とんどがポップスによって構成されていた場合、従来手法を用いると、特徴空間内のポップス領域に楽曲は集中してしまう。上記した偏った特徴空間のイメージ図を図 2 に示す。もしユーザプロフィールが、図 2 における  $a$  や  $b$  のような密度の低い位置に存在しているとしたら、システムがユーザの好みに沿うと思われる楽曲を探し当てるのは容易であり、ほかの楽曲と特徴空間内での区別がつけやすい。それらユーザプロフィール周辺を探索すると、それに見合った検索結果が返ってくるということは容易に想像がつく。しかし、 $x$  や  $y$  などのような、密度の高い位置に存在していた場合、システムが正確にユーザの好みに合った音楽検索を行うことが困難になり、このようにユーザプロフィール周辺に密度が集中した場合、それぞれの楽曲を区別することも難しくなる。

#### 4. 特徴空間再構築手法

上記した特徴空間の分布の偏りは、ジャンル情報に基づいた特徴空間を用いて楽曲ベクトルを生成することで起こりうるため、検索対象データの中から学習データを選別し特徴空間を再構築することで解消できると考えられる。検索対象データ群から学習データを抽出することで検索対象データの特徴を分布させるために最適な特徴空間を構築することができ、ジャンル情報に依存する必要もなくなる。また検索対象データがジャンル情報に対して、どのような分布を持とうとも、対象 DB に最適な特徴空間を生成できる。従来の特徴空間を再構築する手法として、以下の手法を提案する(図 3)。まず、従来手法と同様に学習データに RWC 研究用音楽 DB[3]のジャンル情報  $\{G_1, G_2, \dots, G_m\}$  を用いてツリーを生成する。それに基づいて全検

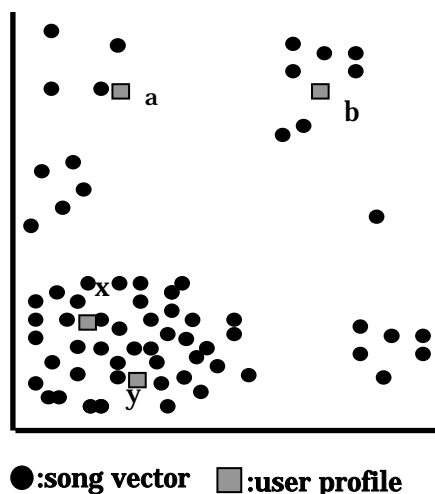


図 2 偏りのある特徴空間イメージ図

索対象データのベクトル量子化を行う。全楽曲ベクトルに対しクラスタリングを行い、クラスタ  $\{C_1, C_2, \dots, C_n\}$  に分割する。その後、各クラスタ重心に対する類似度が高い楽曲を TreeQ の学習データとして抽出し、再度全検索対象データのベクトル化を行うことにより、特徴空間を再構築する。そして、再構築された特徴空間を利用して検索を行う。

#### 5. 実験

この項では本手法を評価するために従来手法と本手法の検索精度の比較を行う。

##### 5.1 実験データ

検索に用いた楽曲は市販されている CD アルバム 120 枚中の楽曲データおよび、コロンビア大学より提供されている us pop data set[4] を元に収集した 6863 曲で構成されている。また、ユーザの嗜好情報として被験者 19 名に従来システムを利用してもらい、それぞれが試聴した楽曲に対して 5 段階の主観評価(好き:5~嫌い:1)を行ってもらった。そしてユーザの評価に基づき楽曲データを 3つのカテゴリ  $C_g, C_f, C_b$  に分類する。各カテゴリの対象ユーザの評価値及び、割合は以下のようにになっている(表 1)。

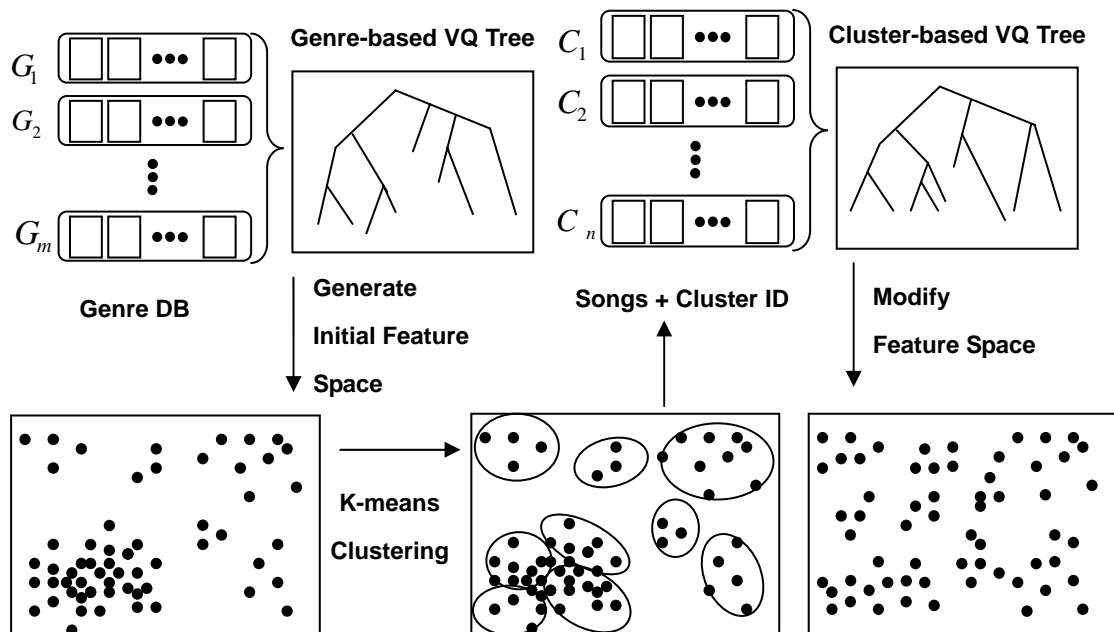


図3 提案手法概要

表1 ユーザ評価値割合

カテゴリ	評価値	割合 (%)
$C_g$	5	15.2
	4	19.5
$C_f$	3	32.1
$C_b$	2	21.8
	1	11.5

### 5.2 偏った特徴空間の分析

上記した特徴空間内の分布の偏りが実際に解消されているのか検証を行った。実験方法としてはまず、従来手法と K-means 法を用いた本手法(以降 KM 手法)の両方で特徴空間の生成を行う。従来手法においては、構築された特徴空間内において、学習データに用いたジャンル情報を元にジャンルプロフィールを作成する。また KM 手法においても同様にクラスタプロフィールを作成する。それらプロフィールと全検索対象データの類似度を計算させ、全楽曲がどのジャンルもしくはクラスタに属するか調査を行った。結果を図4に示す。従来手法では34ジャンル

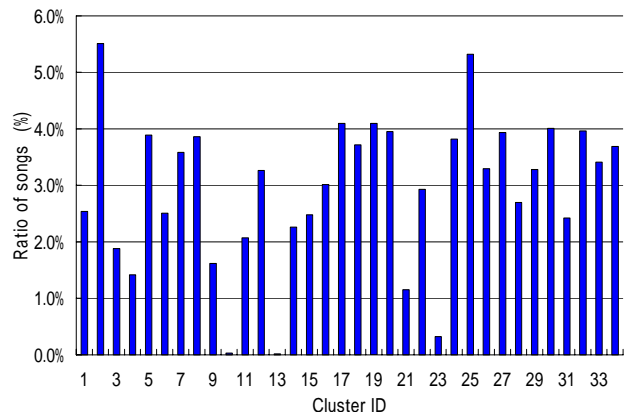


図4 各クラスタに属する楽曲分布

の楽曲を学習データとしているため、公平化を図るために KM 手法においても  $K = 34$  として実験を行った。従来手法では 70.3% の楽曲がポップスに属していることが得られ、特徴空間内において偏りが生じていることが確認できた。本手法では、図4でも確認できるように、1つのクラスタに属する楽曲は最大でも 6%に満たないことがわかり、特徴空間の偏りを解消することができたといえる。

### 5.3 検索精度比較実験

この項では従来手法と本手法の検索精度の比較を行い、本手法の有効性を示す。まず擬似的にユーザプロフィールを作成する。各ユーザのカテゴリ  $C_g$  に属する楽曲からランダムに 5 曲を選び、それに基づいて各ユーザプロフィールを作成する。それらユーザプロフィールと全検索対象データの類似度を計算し、上位 50 曲を出力させる。その類似度上位 50 曲中に各カテゴリに属する楽曲が出現する頻度を調査した。ここで、ランダム性を考慮し、一人の嗜好情報からそれぞれ 5 つのプロフィールを作成し、19 名それぞれ一つずつプロフィールを選択したものを 1 セットとみなし 5 セットそれぞれ試行を行った。各ユーザプロフィールセットにおける検索精度の平均値、最大値及び最小値の結果を図 5 に示す。この図においてクラスタ数 1 の数値は従来手法における検索精度となっている。最大検索精度を示したのは  $K = 15$  における 55.2% で、図からも明らかのようにクラスタ数 4 から検索精度は従来手法の 45.9% を上回り提案手法の有効性を示すことができたといえる。

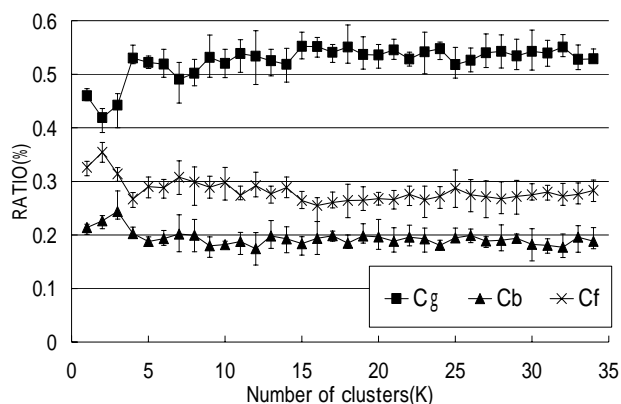


図5 クラスタ数の変動に対する各カテゴリ楽曲出現頻度の平均、最大値及び最小値

### 5.4 初期値変動実験

5.3 節の実験で用いた K-means 法における既知の問題として、初期値選定問題があるが、高次元データではクラスタリング結果の妥当性を確認することが困難であるため、この問題における検索精度に対する影響を調査した。また比較対象として K-means 法の拡張であり初期値にデータ全体の重心を用いることで初期値問題への影響が少ない LBG アルゴリズム[5]に基づいたクラスタリング(以降 LBG 手法)から KM 手法と同様に学習データを抽出し、検索精度を出力する。KM 手法では初期値を変動させ、5.3 節の実験と同様に検索精度を出力する。100 回の試行から各試行における検索精度  $F(C_g)$  を出力し、出現頻度の分布を図 6 に示す。なお、クラスタ数は 5.3 節の実験において最も精度の良い  $K = 15$  を用いた。LBG 法についてもクラスタ分割における閾値の調整を行い、クラスタ数を 15 とした。また KM 手法 100 回の平均 LBG 手法の検索精度を表 2 に示す。KM 手法における検索精度の最大値は 57.2%、最小値が 38.9% となり図 6 から明らかに初期値の違いから検索精度に大きな差が見られた。

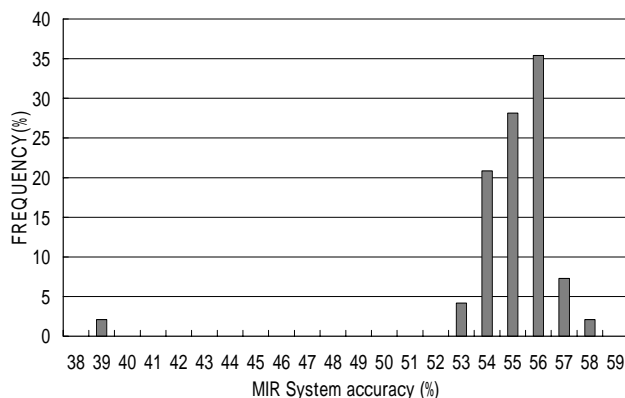


図6 KM 手法における検索精度分布

表2 各手法検索精度結果 ( $F(C_g)$ )

KM 手法	LBG 手法
54.5%	54.8%

図6の約39%の検索精度となったのは特徴空間内でのクラスタリング結果が本来の楽曲の特徴分布に準じていなかったためと考えられ、実際にシステムに適用するにはこの精度の差は明らかに問題である。一方、LBG手法においてはKM手法における試行の平均値よりも高い精度を示し安定して高い検索精度が得られることが判明した。システム利用を考慮した場合、高次元データに対してはクラスタリング結果の妥当性を検証するのは困難であり、実験のように検索精度から初期値の正誤を決められるわけではないため、LBG手法のような安定して有効な検索精度を得られるクラスタリング手法が望まれる。しかし、KM手法において平均値は54.5%であったが、LBG手法の精度を越える事例は72を超えていた。KM手法の検索精度の劣化を深く検証し、その問題を改善することができれば安定した検索精度の結果を得られる可能性はある。クラスタ分類能力においてはLBG法よりもK-means法の方が優秀であり、本来の楽曲の特徴に準じたクラスタリングを基に学習データを抽出することが最適な特徴空間の構築につながる。そのため、KM手法の改善を行い、クラスタリング結果の安定性を確保する事で、検索精度が向上できると考えられる。

## 6. まとめ

本稿ではユーザ嗜好に基づく音楽情報検索システムにおける特徴空間内の楽曲に対しクラスタリングを行うことで学習データ

を抽出し、特徴空間の再構築を行うことで特徴空間の偏りが改善されることを確認し、本手法によって検索精度が向上することを示した。またKM手法における初期値選定問題についてLBG手法との検索精度の比較を行うことで、両手法のシステム適用に関して考察を行い、未だ改善の余地はあるものの分類能力の高さからKM手法の方がシステムに適していると考えられる。今後、その他の手法との比較を行い、システムに最適なクラスタリング手法の検討を行う必要がある。

## 7. 参考文献

- [1]K.Hoashi,K.Matsumoto,N.Inoue:“Personalization of user profiles for content-based music retrieval based on relevance feedback”, Proceedings of ACM Multimedia 2003,pp 110-119,2003
- [2]J.Foote:“Content-based retrieval of music and audio”, Proceedings of SPIE, Vol 3229,pp 138-147,1997
- [3]M.Goto,H.Hashiguchi,T.Nishimura,R.Okazaki: “RWC Music Database: Music Genre Database and Musical Instrument Sound Database.”, Proc. of the 4th Int'l Conference on Music Information Retrieval (ISMIR 2003), pp.229-230,2003.
- [4]D.Ellis:“The uspop2002 Pop Music data set”, List available at <http://www.ee.columbia.edu/%7Edpw/research/musicsim/uspop.html>, 2003
- [5]Y. Linde, A. Buzo, and R. M. Gray:” An algorithm for vector quantization design”, IEEE Trans. on Communications, Vol. COM-28, 84-95,1980