

音楽情報検索システムの操作履歴とユーザ嗜好の相関分析

安念 高志[†] 帆足 啓一郎[‡] 松本 一則[‡] 甲藤 二郎[†]早稲田大学理工学部[†] KDDI 研究所[‡]

1. はじめに

ユーザ嗜好に基づいた音楽情報検索システムにおいて、適合フィードバック情報を利用することにより、その精度が上がることはすでに示されている.[1]しかし、ユーザが個々の楽曲に対し、適合フィードバック情報を与えるのは負担が大きい。そこで、ユーザのシステム操作履歴を、暗示的フィードバックとして利用することにより、その負担を軽減することを検討する。

本稿では、システム操作履歴とユーザの楽曲嗜好の関係を調べるため、実際にシステムを使用したの評価実験を行い、ユーザが楽曲を試聴する際に、どのようなアクションを起こしたかを操作ログとして保存し、そのログがユーザの楽曲への嗜好に対し、どのような関係を持っているのか、相関分析によって割り出した。

2. 従来の手法と課題

2.1 音楽情報検索システム

本システムではユーザの音楽嗜好を基に、適合フィードバックを行うことにより、ユーザ嗜好に沿った楽曲の検索を行う。

まずツリーベクトル量子化手法(TreeQ)[2]を基に、全検索対象楽曲についてベクトル量子化を行う。また、ユーザの好みに沿ったジャンルに基づき、ユーザ嗜好ベクトルを作成する。このユーザ嗜好ベクトルと検索対象データベクトルの中から、類似度の高いものを検索結果として提示する。

ユーザは提示された楽曲に対し、評価を行い、それを元にユーザ嗜好ベクトルを更新し、再度類似度の高い検索対象楽曲を提示する。こうして適合フィードバックを繰り返し与えることにより、より検索の精度を向上させることができる。

2.2 従来手法の課題

ユーザ嗜好に沿った楽曲を、より高い精度で検索するためには、ユーザが楽曲ごとに評価(適合フィードバック情報)をシステムに与え、検索を繰り返し行う必要がある。楽曲を試聴しつつ、その都度楽曲の評価を行うのは、ユーザにとって負担になりうる。

3. 提案手法

上記の課題をクリアするために、ユーザが楽曲の試聴中に行ったシステム操作履歴の暗示的フィードバックを利用する。システム操作履歴とは、ユーザが楽曲を試聴する際にシステムに対し行った操作、具体的には再生や停止などである。これらのシステム操作履歴(ログ)が、ユーザの楽曲に対する嗜好と何らかの相関性を持っているとすれば、それらの操作が行なわれる度に、システムに適当な評価データとしてフィードバックすることにより、ユーザに負担をかけることなく有効なフィードバック情報を得ることが可能となる。つまり、それぞれのログの嗜好との相関に応じて、フィードバックの重みを調整し、評価データとしてシステムに与えることで、ユーザの楽曲評価の代役を果たすようにする。

4. 実験

暗示的フィードバックを利用するためには、システム操作履歴とユーザの嗜好との相関が不可欠である。そこで相関の有無を調べるため、以下のような実験を行った。

4.1 実験概要

期間約一ヶ月、被験者 28 名に対し、実際に音楽情報検索システムを使って評価実験を行った。実験システムでは、はじめに被験者はランダムに提示された楽曲に対し、「好き」「嫌い」の評価を行う。システムはその評価に基づいて、ユーザの嗜好ベクトルを作成し、ユーザの嗜好に合うベクトルを検索し、それを提示する。再度提示された楽曲に対し、ユーザはさらに評価を行い、ユーザの嗜好ベクトルを更新していく。こうして、ユーザ嗜好に沿ったプロファイルを作成する。

実験期間中に、被験者がインターフェース上

Correlation of user Preferences and operation log of music information retrieval system

[†]Takashi Annen [‡]Keiichiro Hoashi [‡]Kazunori Matsumoto [†]Jiro Katto

[†]Graduate school of Science & Engineering, Waseda University [‡]KDDI R&D Laboratories Inc.

(図1)で行った操作はログは相関分析のために収集した。対象としたログは表1示す全8種類である。



図1 実験システムインターフェース

表1 相関分析対象のログの種類

AutoPlay	楽曲の自動再生
Good	ユーザが「好き」をチェック
Bad	ユーザが「嫌い」をチェック
Select	検索結果から楽曲を選択
DoubleClick	検索結果から選択, または再生
Skip	再生中の曲を停止し, 次曲を再生
Stop	ユーザが曲を停止
AutoListen	1曲の再生が最後まで終了

また, 被験者には評価実験の最後に, 試聴してもらった楽曲について, 5段階(1が「嫌い」、5が「好き」)での評価をつけてもらった。これらのデータから二乗検定を用いて相関分析を行った。

4.2 実験結果

評価実験の結果集まった81,469行のログと, 15,423曲の嗜好の評価データに対し, 二乗検定を行った結果は表2のようになった。

表2 各ログとユーザ嗜好の二乗検定結果

	値	相関の有無
AutoPlay	1	無
Good	0.001	有
Bad	0.001	有
Select	1	無
DoubleClick	0.2	無
Skip	0.001	有
Stop	0.001	有
AutoListen	0.001	有

二乗検定の結果では, 相関を持つログは Good, Bad, Stop, Skip, AutoListen となっている。Good と Bad については, ユーザ嗜好の明示的フィードバックなので, 相関を持つのは当然であるが, Skip, Stop, AutoListen が Good や Bad に劣らない相関を示しており, これらについては暗示的フィードバックを行うに十分な統計的有意差があると認められる。一方, 相関の弱い AutoPlay, Select, DoubleClick については, ユーザが楽曲を試聴する直前に行っている操作であるため, ユーザ嗜好との相関が弱かったと予想される。

また, 相関を持つログが, 「好き」「嫌い」どちらの嗜好と相関を持つのか調べるため, それぞれの対象楽曲の評価値平均を求めた。(表3)

表3 各ログの対象曲の評価平均

	平均評価値
Good	4.357
Bad	2.316
Skip	2.702
Stop	3.691
AutoListen	3.300

以上より, Stop, AutoListen はユーザ嗜好に沿った評価と, Skip はユーザ嗜好に沿わない評価と相関を持っていることがわかった。

5. まとめ

本研究では, 音楽情報検索システムに, ユーザのシステム操作履歴を暗示的フィードバックとして利用することにより, ユーザの負担を減らし, 検索精度を向上させるため, システム操作履歴とユーザ嗜好の相関分析を行った。

評価実験で収集したデータの相関分析の結果, 一部のシステム操作履歴と, 検索された楽曲に対する評価には強い相関があることが確認された。この結果をもとに, 今後, 相関の強弱による暗示的フィードバックのパラメータの最適化を行い, 音楽情報検索精度の向上を達成するアルゴリズムを実装し, その評価実験を行っていく。

参考文献

- [1]Hoashi,Matsumoto,Inoue:Personalization of User Profiles for Content-based Music Retrieval based on Relevance feedback,Proceedings of ACM Multimedia2003,Oct,2003
- [2]Foote:"Content-based retrieval of music and audio",Proceedings of SPIE,Vol 3229,pp138-147,1997