

自動作曲システムにおける協調フィルタリングを用いた 楽曲評価予測

Prediction of Evaluation for Automatic Tune Composition Based on Collaborative Filtering

*1 杉本知仁 *1 太田晶大 *2 森山甲一 *2 栗原聡 *2 沼尾正行
Toshihito Sugimoto Akihiro Ota Koichi Moriyama Satoshi Kurihara Masayuki Numao

*1 大阪大学大学院情報科学研究科情報数学専攻

Department of Information and Physical Sciences, Graduate School of Informartion Science and Technology, Osaka University

*2 大阪大学産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

We aim to propose efficient algorithms and background knowledge extraction techniques for machine learning. One of our unique research problems is to develop an automatic music composition system that is tightly coupled with the user's impressions of music. The system learns emotion from user's evaluations of tunes on the assumption that their evaluations reflect their own sensibility. However, at the stage of user's evaluation, it took long time to evaluate music. Therefore This paper proposes a prediction technique of evaluation based on Collaborative Filtering which reduces the time and user's burden for evaluation. Sensitivity is different by a person, but it is a fact that there is a person having similar sensitivity. So, we thought that people with similar sensitivity are be able to divided into the same sensitivity group.

1. 序論

コンピュータの誕生以来、情報技術の著しい発展は社会基盤を大きく変化させてきた。しかし、これまでの情報システムでは、一人ひとりの異なる考え方、嗜好、知識の違いなどは考慮されることがほとんどなかった。さまざまな違いを持つコンピュータ利用者が便利かつ快適に生活するためには、情報システムが個人の心つまり感性を理解する必要がある。このような背景から近年、感性を扱う研究が盛んに行われている。音楽に関する分野でも、音楽が人間の感性に与える影響についての研究や、人間の感性を用いた音楽検索等、感性に関連する幅広い研究が行われてきている [1]。本研究室では感性処理の研究の一環として、人間の感性を学習し、得られた知識を用いて個人の嗜好に沿った楽曲を自動生成するシステム [2, 3] を提案・構築してきた。

まず、感性を学習する為にユーザに既存の楽曲を聴かせて、それに対する評価値を得る。次に、楽曲の情報とユーザの評価値から、ユーザの感性と関連のある楽曲構造を、帰納論理プログラミング (ILP) を用いて一階述語論理形式で学習する。そして、得られた感性モデルを基に適合度関数を構成し、遺伝的アルゴリズム (GA) によって作曲を行う。

従来の研究では、個人の感性を個別に扱ってきた。しかし、感性は人によって様々に異なるが、似たような感性を持つ人がいることも考えられる。そこで本研究は、音楽を聴いたときのユーザの感性を過去のデータから予測することを提案する。

ユーザにある程度の曲数を評価してもらい、評価していない曲は協調フィルタリング手法を用いて自動的に評価付けを行う。これにより訓練曲を評価するユーザの負担が減り、作曲までの時間を短縮することが可能になる。

協調フィルタリングは Amazon.com の「おすすめの商品」

や、はてなアンテナの「おとなりアンテナ」など主に推薦システムに応用されている。ここでは同手法を用いてユーザ間類似度を算出することにより学習に用いる訓練例の評価付けに利用した。また、訓練曲間の類似度を計算することにより、類似した曲の評価を避け、予測精度の向上を目指した。概要図を図 1 に示す。

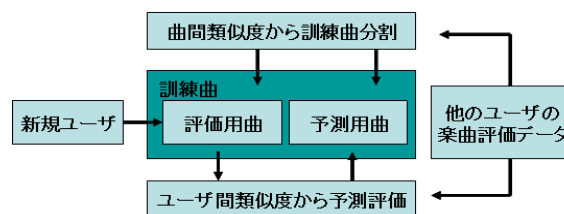


図 1: 概要図

2. 先行研究と関連研究

2.1 論文の位置付け

本研究はユーザフレンドリーな自動作曲システムの構築のための一研究である。従来のシステムでは訓練曲評価段階でユーザに大きな負担がかかってしまうことが問題となっており、これを解消することを目的としている。

2.2 先行研究

従来システムでは、感性語と呼ばれる形容詞にアンケート方式で 5 段階評価を付ける SD 法により楽曲に対する評価値を得ていた。しかし、SD 法の再現性の保持の難しさや評価判断基準の変化し易さといった疑問点を解消するため、生理-生体信号である脳波の感性スペクトル解析手法 [4] を用いた楽曲評価手法が提案された。この手法を用いると、音楽提示時のユーザの感情が連続データとして得られ、曲中での感情の変化を捉

連絡先: 杉本知仁, 大阪大学, 産業科学研究所沼尾研究室,

〒 567-0047 大阪府茨木市美穂ヶ丘 8-1,

Tel:06-6879-8426 Fax:06-6879-8428

E-mail:sugimoto@ai.sanken.osaka-u.ac.jp

えることが可能となる．そこで楽曲を1小節ごとに評価することによって訓練曲数を減らす手法が試みられたが，小節単位の評価では楽曲単位で決まる調やテンポのような属性についての情報が少なくなるという欠点があった．

2.3 関連研究

協調フィルタリングとは，多くのユーザの嗜好情報を蓄積し，あるユーザと嗜好の類似した他のユーザの情報を用いて自動的に推論を行う方法論であり，リコメンデーションやパーソナライゼーションに利用される代表的な手法である．これはキーワードと内容の比較によるフィルタリングである *Content-based Filtering* や，関心の表現に得点を用いるフィルタリングである *Rating-based Filtering* などに分類できる．本研究では，音楽のような内容の分析が難しい研究に有効である評価値予測方法として *Rating-based Filtering* を用いる．

この分野の代表的な研究である GroupLens[5] はネットニュースにおけるフィルタリングシステムで，ユーザ間の類似度を計算し未読のニュースに対する評価値を予測することによってユーザがより高い評価をつけるであろうニュースを推薦する．また，Ringo[6] はユーザが音楽 CD に対して評価付けを行い，蓄積した嗜好情報と利用者ユーザとの嗜好情報の比較から推薦を行う．これは評価軸の中心値を基準点とし，その値より大きい値はその対象に対して肯定的な評価，小さい値はその対象に対して否定的な評価と分けて相関係数を算出する．

このように協調フィルタリングではユーザ間類似度を定義することによって個人差を考慮した評価値予測ができる．本研究では，GroupLens の類似度計算式，評価値予測式を参考にし，Ringo の基準点の考え方を応用している．

3. 提案手法

本研究における協調フィルタリングを用いた評価値予測手法の概念図を図2に示す．

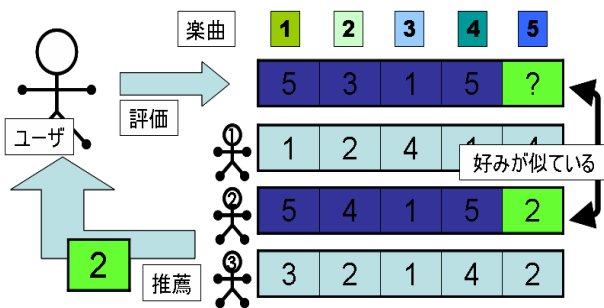


図2: 概念図

3.1 評価値予測手法

まず全訓練曲を評価用曲セットと予測用曲セットの2つに分割し，新ユーザに評価用曲セットを評価してもらう．その評価データと他の全てのユーザの全訓練曲評価データを用いて新ユーザの予測用曲セットの評価値予測を行う．評価値予測には GroupLens の類似度計算式，評価値予測式を参考にした．

まず，類似度計算式について説明する．(1)式はユーザ A とユーザ B の類似度を示している．

$$\gamma_{AB} = \frac{\sum_i (A_i - X)(B_i - X)}{\sqrt{\sum_i (A_i - X)^2} \sqrt{\sum_i (B_i - X)^2}} \quad (1)$$

A_i, B_i はそれぞれユーザ A, B の i 曲目評価値を示している． X はユーザ毎の平均評価値を適用するが，本研究ではユーザの評価値が基準値より正方向か負方向であるかが重要であるため X に評価値の基準値を適用した．つまり，5段階評価であれば $X=3$ になる．

次に，評価値予測式について説明する．(2)式は「ユーザ A の i 曲目の評価値」を予測する算出式である．

$$A_i = X + \frac{\sum_{B \in \text{raters}} (B_i - X) \gamma_{AB}}{\sum_B |\gamma_{AB}|} \quad (2)$$

評価値予測式は (1) 式で算出した類似度で他者の評価値に重み付けを行うことにより，他者とのずれの度合いを算出するものである．従って，予測評価値は類似度が高いユーザの評価値には大きく影響され，類似度が低いユーザの評価値にはほとんど影響されない．

つまり評価値予測とは，新ユーザと全過去ユーザの評価用曲セットの評価値を基にユーザ間類似度をそれぞれ算出し，その類似度と全過去ユーザの予測用曲セットの評価値を用いて，新ユーザの予測用曲セットの評価値を予測することである．

3.2 評価用曲選択手法

訓練曲を評価用曲セットと予測用曲セットに分割する方法が評価値予測の精度に依存すると仮定し，評価用曲の選択を適切に行う必要があると考えた．そこで，3.1節で説明した類似度計算式をユーザ間から曲間に応用することによって，他曲との関連性が高い曲を評価用曲として選択することを提案した．以下に曲間類似度を用いた評価用曲選択アルゴリズムを示す．

```

begin
  評価用曲リスト PL を初期化
  予測用曲リスト TL を初期化
  while PL < 定めた評価用曲数 do
    訓練曲リスト ML より他曲に対する類似度リスト SL
    を算出
    全ての SL で最も類似度が高い曲を PL に格納
    格納した曲を ML から削除
  end while
  ML に残った曲を TL に格納
end
    
```

類似度リスト SL を作成する式を (3) に示す． K, L には全訓練曲リスト ML の中に含まれる楽曲の全てのペアが適用される．

$$\begin{aligned} \gamma_{KL} &= \frac{\text{Cov}(K, L)}{\sigma_K \sigma_L} \\ &= \frac{\sum_i (K_i - \bar{K})(L_i - \bar{L})}{\sqrt{\sum_i (K_i - \bar{K})^2} \sqrt{\sum_i (L_i - \bar{L})^2}} \quad (3) \end{aligned}$$

4. 実験

被験者実験によってユーザの楽曲評価データを獲得し、提案した手法の有用性を検証する。今回は Ringo の欠点である嗜好の解釈のずれを解消するため、SD 法ではなく脳波解析によって訓練曲評価を行った。提示した訓練曲は 75 曲、ユーザは 15 人であり、評価される感性は stress, joy, sad, relax の 4 種類である。

4.1 評価曲数と予測精度

提案した手法の評価用曲数に対する予測精度を検証するための実験を行い、その結果を図 3 に示す。実験は 14 人のデータを用いて 1 人のテスト曲を予測する *Leave-one-out* 方式で行い、評価用曲数を 10, 20, 30, 40, 50 曲とした場合の予測評価曲の平均正答率を比較した。

ここではユーザ本人の予測用曲の実際の評価値と予測評価値の方向性（基準値より大きいか小さいか）が一致した例を正例とし、一致しない例を負例とする。また、正答率は全ユーザに対して評価値予測を行った正例の割合である。

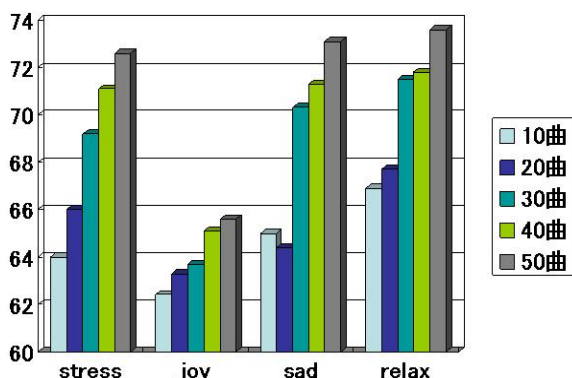


図 3: 評価用曲数の変化と予測正答率

図 3 より、類似度算出に用いられる評価用曲数が多い方が予測正答率が高いことが分かる。これより、評価値予測が適切に行われていることが確認できた。

4.2 評価用曲選択と予測精度

評価用曲の選択が予測精度にどの程度影響を与えるかを検証する実験を行った。まず評価用曲を 30 曲ランダムに選択し、14 人のデータから 1 人のテスト曲を予測する作業を全ユーザに対して行う。この作業を 10 セット行い、平均正答率と標準偏差を感性別に算出したものを表 1 に示す。

次に 15 人・75 曲の評価値データを基に 3.2 章で述べた曲間の類似度を用いての評価用曲選択を行う。選択された評価用曲を用いて、14 人のデータから 1 人の予測用曲を予測する作業を全ユーザに対して行った。評価用曲の選択は感性毎に行われる。類似度が高い曲、低い曲をそれぞれ評価用曲として選択した結果を表 2,3 に示す。類似度が低い曲の選択は比較のために行っている。

また、上記 3 種類の評価用曲選択方法における平均正答率を比較し、図 4 に示す。

表 1: ランダムに評価用曲選択を行った場合の正答率

感性	平均正答率 (正例/負例)	標準偏差
stress	69.2% (4671/2079)	2.6%
joy	63.7% (4299/2451)	2.4%
sad	70.3% (4745/2005)	1.6%
relax	71.5% (4826/1924)	1.2%
全感性平均	68.7%	1.9%

表 2: 曲間類似度 (高) を用いた場合の予測正答率

感性	類似度 (高) 平均正答率 (正例/負例)	標準偏差
stress	73.5% (496/179)	4.8%
joy	71.0% (479/196)	4.3%
sad	74.4% (502/173)	5.0%
relax	76.0% (513/162)	5.7%
全感性平均	73.5%	5.0%

表 3: 曲間類似度 (低) を用いた場合の予測正答率

感性	類似度 (低) 平均正答率 (正例/負例)	標準偏差
stress	71.5% (483/192)	5.1%
joy	60.0% (405/270)	3.0%
sad	70.2% (474/201)	4.9%
relax	70.3% (475/200)	6.1%
全感性平均	68.0%	4.7%

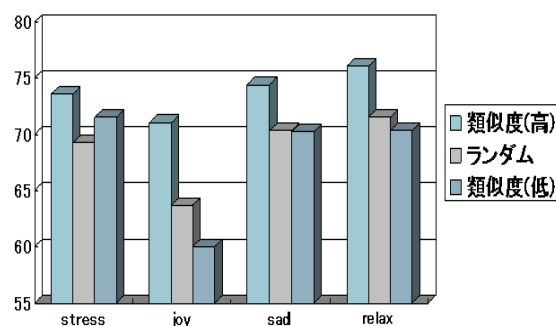


図 4: 評価用曲数の選択法と予測正答率

図 4 より、感性 stress については一部逆転が見られるが、類似度の高い曲を選択した場合が最も正答率が高く、次いでランダム選択、類似度が低い曲を選択した場合という結果となった。これより、類似度を用いた評価用曲選択は十分効果を発揮していると思われる。

次に実際に図4から得られる曲間類似度(高)と曲間類似度(低)の予測正答率の差が有意なものであるか確かめるためにt検定を行った。結果を表4に示す。

表4: 検定結果

	stress	joy	sad	relax
分散	0.29	0.89	0.07	0.35
t 値	0.70	2.95	1.21	2.01
有意差 (5%)	なし	あり	なし	あり
有意差 (1%)	なし	あり	なし	なし

表4より joy, relax に関してはその差が有意であることが確認された。

5. 考察

評価用楽曲をランダムに選択した場合は68.7%の精度で、曲間類似度を用いて適切に評価用曲を選択した場合は73.5%の精度で評価値予測を行うことができた。検定の結果、joy, relaxの2感性に関して評価用曲選択が有意であることが確認されたが、これは類似度の高い楽曲が多かったためと考えられる。今後、評価用曲選択手法をさらに改善することにより、より高い精度での予測が可能であると考えられる。

6. 議論

本研究では *Rating-based Filtering* を用いた。今後は、音楽情報のコンテンツ情報も考慮した *Content-based Filtering* を適用したより高度な手法も検討する必要がある。

提案した類似度を用いた評価用曲選択手法では、感性ごとに独立して評価用曲選択を行う。従って全感性について作曲しようとした場合、結局75曲全てに対して評価することになりかねない。そこで類似度の計算は感性ごとに独立に行ってそれぞれから候補を出し、重複があればなるべくそれを採用できるようなアルゴリズムを考える必要がある。

また、評価値予測によって得られる評価値と実際に得られる評価値を同等に扱ってはいけなくと考えられるため、学習の際に何らかの重み付けを行う必要がある。

7. 結論

自動作曲システムにおいて協調フィルタリングを用いた楽曲評価値予測手法を提案した。曲間類似度を用いて評価用曲を適切に選択し、ユーザ間類似度を用いて楽曲の評価値予測を行うことが有意義であることを示した。また本手法を用いることにより、楽曲評価時のユーザの負担を軽減させることが可能であることを示した。

参考文献

- [1] 長島知正, 久保洋, 魚住超, 金木則明. 感性と情報. 森北出版株式会社, 2007.
- [2] Roberto Legaspi, Yuya Hashimoto, Koichi Moriyama, Satoshi Kurihara and Masayuki Numao. Music Compositional Intelligence with an Affective Flavor. *Proc. ACM International Conference on Intelligent User Interfaces*, pp.216-224, 2007.

- [3] Toshihito Sugimoto, Roberto Legaspi, Akihiro Ota, Koichi Moriyama, Satoshi Kurihara and Masayuki Numao. Modelling Affective-based Music Compositional Intelligence with the Aid of ANS Analyses. *Proc. 27th SGAI International Conference on Artificial Intelligence*, 2007.
- [4] Toshimitsu Musha and Yuniko Terasaki. Feature extraction from EEGs associated with emotions. *Artificial Life and Robotics*, Vol.1, pp. 15-19, 1997.
- [5] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom and John Riedl. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. *In Proceedings of the 1994 Computer Supported Collaborative Work Conference.*, pp.175-186, 1994.
- [6] Shardanand U and Pattie Maes. Social Information Filtering: Algorithm for Automating " Word of Mouth. *ICHI '95 Conference Proceedings.*, pp.210-217, 1995.