

利用者の印象に基づく音楽レコメンドサービス： 画像を利用した利用者指向の音楽レコメンド手法の考察

中野美由紀, 石山悟志, 君野史明, 小久保勇氣, 斉藤大介, 鈴木智尚, 宮崎晃一, 吉村慎祐[†]
産業技術大学院大学[†]

1. はじめに

画像, 音楽など多くのコンテンツがインターネット上にあり, ユーザにとって有益な情報を得るために様々なサービスが検討されている. 特にユーザの嗜好が的確に表現しにくい音楽, 画像, 動画に関する推薦サービスでは, ユーザの好みを反映するために精緻なかつ多数の入力情報を求めることが多く, 同じ様なコンテンツを選ぶ他のユーザの情報を利用するなどの工夫がなされている. 一方で, 気軽に楽しみたいユーザにとっては簡易な入力, ユーザにとって興味深い結果が得られることが望ましい. 本サービスでは, 直感的かつ簡易な入力として画像を利用し, 画像の印象に適した音楽を推薦することを目指している. 印象という曖昧な条件を介在することで, 利用者にとって, 希望する結果に加え, 従来は聞いたことのない音楽, 思いもかけない結果を提示することが期待できる.

本論文では, 音楽検索条件として画像選択インタフェースを用い, 音楽を聴きたい利用者のそのときの気分に合わせて音楽を推薦するサービス”Diggin”¹⁾を開発した. 当該サービスでは, 従来の具体的な検索条件ではなく感情空間²⁾を基にした推薦を行うため, 従来は気づけなかった新しい音楽との出会い(セレンデピティ)の場をも提供する. 音楽データベースとして Spotify を中心に収集した音楽コンテンツ情報を用い, 感情空間上に音楽データベースのマッピングについて検討する. さらに, 音楽, 画像の推薦で感情空間を利用した手法が近年提案されており^{3,4,5)}, ”Diggin”では, 5)で提案された画像からの感情空間へのマッピング手法を利用し, ”Diggin”音楽データベースから適切な楽曲を推薦する. 本稿では, 利用者の入力インタフェースとして開発した画像選択と感情空間の検討について報告する.

2. 音楽推薦サービス: Diggin¹⁾

”Diggin”では公開されている音楽情報(楽曲名, 演奏者, 音楽の雰囲気, テンポ, 調等)を Spotify から収集し, 推薦のためのデータベースを構築している. 感情空間と音楽データベースのマッピングは Spotify が提供する楽曲情報(音楽の雰囲気, テンポ等)を用い, 人手を用いた検証を行っている. また, 画像インタフェースとして用いる画像は Flickr から自由に利用可能なものを収集し, また, キーワードを利用した検索も考慮し, GENIUS から歌詞の収集を行った.

3. Diggin 音楽データベース

Spotify が提供している音楽情報は, 楽曲名, 歌手, 作曲家, 編曲者等から曲の調, テンポ等およそ 30 項目近くの情報がある. ポップ, ロック, クラシック等のジャンル別の推薦は現在の音楽配信サービスでもよく用いられているが, 我々が目指す印象に基づいた推薦には適さない. そこで, popularity, energy, valence, tempo 等印象を表すと思われる特徴量で数値化されているものを利用し, 配信される音楽情報の傾向を調べた. 図 1 に四つの特長量と音楽の分布を示す.

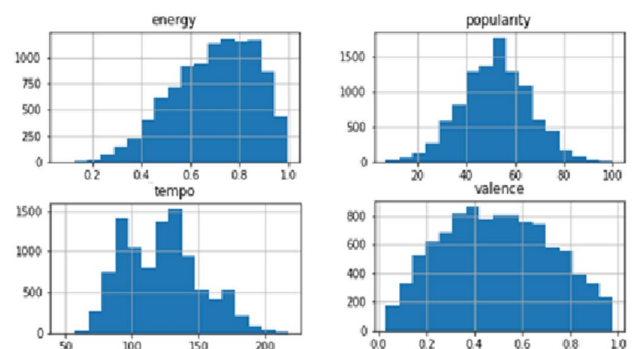


図 1 音楽データベースの楽曲分布

これらの特徴量では, 曲が広く分散しており, 曲の印象との関係づけに利用可能と考えられる. 数値で表現されている特徴量として, Spotify で提供されている 10 個の特徴量について相関関係の在りなしを調べた. 図 2 に示すように, それぞれの特徴量の間は相関関係はさほど強くなく, それぞれ独立の特徴量として利用できることが分かった.

Music Recommendation Service Based on Users' Impression: Consideration on a methodology of user oriented music recommendation utilizing images as an user interface, [†]Miyuki Nakano, Satoshi Ishiyama, Shiro Kimino, Yuki Kokubo, Daisuke Saito, Tomohisa Suzuki, Koichi Miyazaki, Shinsuke Yoshimura · Advanced Institute of Industrial Technology

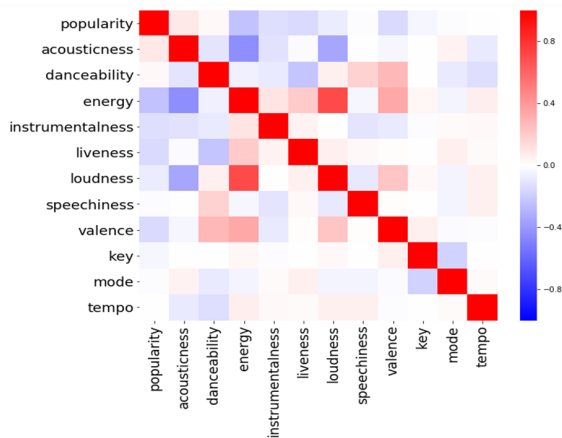


図 2 曲の特徴量の相関関係

感情空間へのマッピングとしてどの特徴量を利用するかについては、異なるサンプル曲（100曲）を選び、印象の異なる曲でどの特徴量が大きく変化しているかについて、人手（7名）により調べた。その結果、人が視聴して大きく変化していると考えられる特徴として、energy と valence が抽出された。表 1 に結果を示す。

表 1 特徴量と曲の印象の関係

artist_name	name	popularity	energy	tempo	valence
The Beatles	Yesterday - Remastered	65	0.179	96.529	0.315
DREAMS COME TRUE	Miraiyosouzu	22	0.479	131.993	0.476
BABYMETAL	Gimme Chocolate!! - live	32	0.998	109.994	0.181
Ludwig van Beethoven	Symphony No. 7 in A Major, Op. 92: II. Allegretto	44	0.0356	65.048	0.0466
Kyary Pamyu Pamyu	ファッション・モンスター	39	0.974	160.009	0.504
Michael Jackson	Beat It	54	0.799	138.825	0.906

Energy と valence はそのまま感情空間の Arousal および Valence とほぼ同じ特徴であることから、Diggin の音楽データベースでは楽曲の感情空間 AV 値として、特徴量 energy と valence の値をベースにマッピングを行うこととした。単純に energy と valence の値を用いたときに、感情空間上にどのようにマッピングされているかを図 3 に示す。

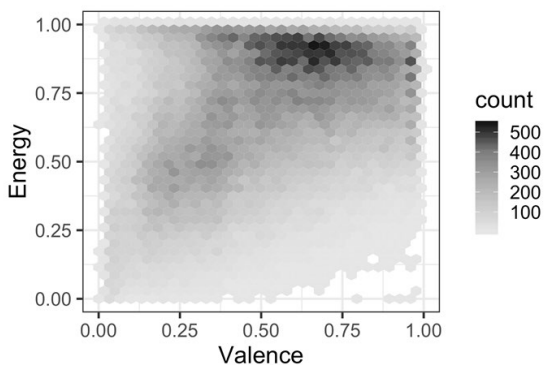


図 4 Energy と Valence における楽曲の分布

図 3 では、収集された 65 万曲の分布を示しているが、右側の象限に多くの曲が固まっている。そこで、人手による曲の視聴を行い、感情空間を四象限に分け、いずれの象限にはいるかについての評価を行った。AV 値としての定量的な値を人の印象として与えることは難しいため、今回は四象限評価としたが、この結果から感情空間を四象限に分けると、中心点はおよそ Energy0.75, Valence0.5 を用いる形で検討を行っている。

4. 画像を用いた推薦アルゴリズム

Diggin では、印象入力インタフェースとして複数の画像から利用者に欲しい曲の印象にあわせた画像を選択させ、その画像の感情空間上の値を用いて近い感情空間上に分布する楽曲の推薦を行っている。画像は Flickr 等から自由に利用できる画像を 3000 枚ほど収集している。また、画像の AV 値は画像の中の物体、背景などを抽出し、得られた名詞等から関連形容詞を得ることで計算している。画像インタフェースイメージを図 4 に示す。



図 3 Diggin 入寮画面と推薦結果

5. まとめ

利用者の印象に合わせた楽曲推薦サービス Diggin について紹介した。今後、利用者アンケート等で推薦内容の精度について検討を行いたい。

参考文献

- 1) <http://nakano03.aait.ac.jp/>
- 2) Russel, J.A., "A circumplex model of affect", *Journal of Personality and Social Psychology*, 39(6), pp.1161-1178, 1980
- 3) 糸井勇貴, 他. "楽曲の風景特徴化に基づく風景ウェア楽曲推薦システム", *DEIM Fotum*, A8-3, 2017
- 4) 新徳龍太郎, 他. "画像の印象に合う楽曲の自動推薦システムに関する研究", *映像情報メディア学会メディア工学研究会技術報告*, ME2013-7, pp.23-26, 2013
- 5) 追木智明, 他. "画像の印象に合った音楽の推薦システムの提案", *DEIM Forum*, P2-3, 2018