

動画特微量からの印象推定に基づく動画 BGM の自動素材選出

清水柚里奈^{†1} 菅野沙也^{†1} 伊藤貴之^{†1} 嵯峨山茂樹^{†2} 高塚正浩^{†3}

概要: 動画を撮影して SNS で公開する際に、BGM を付与して楽しむ人が増え、またそれを支援するアプリも増えてきた。本研究では動きや色、被写体のキーワードといった動画特微量から印象を推定し、その結果に基づいて選出されたメロディとリズムをマッシュアップする楽曲生成を行うことで、動画の印象に合った楽曲を付与する手法を提案する。まず動画から一定時間ごとに色および動きの特微量を算出し、それに基づいて動画の印象を推定する。また予め用意したメロディとリズム進行について被験者に印象を回答してもらい、動画の印象値と最も類似度の高いものを選びだし、それらを合成する。さらにコード進行を付与し、反復回数などを調整することで、動画の長さに合った楽曲を生成する。以上の処理により、印象に合った音楽を自分で探すことなく動画に付与することができる。

キーワード: 動画解析, 色, 動き, 楽曲生成, 音楽特微量, 印象推定

Automatic Background Music Composition Based On Impression Estimation of Input Movies

Yurina Shimizu^{†1} Saya Kanno^{†1} Takayuki Itoh^{†1}
Shigeki Sagayama^{†2} Masahiro Takatsuka^{†3}

Abstract: Recently many people enjoy accompanying background music to the movies in uploading movies in social Web services. Many applications and services to assist the background music editing have been released. This paper presents a technique to automatically generate the background music that matches impression of movies. The technique estimates impression of movies from the feature values of movement and color. It then generates the background music by synthesizing melody and rhythm selected based on the impression. The technique learns the relationship between features of movies or music and their impressions answered by the users, so that the music generation process reflects the users' own impression. Users can accompany preferable background music to the movies by this technique, without searching for the tunes by themselves.

Keywords: movie analysis, color, movement, music generation, acoustic feature, impression estimation

1. はじめに

近年、デジタルカメラやスマートフォンの普及により、写真や動画を撮影する機会が増え、またその撮影したものを SNS サイトに投稿することで、多くの人々と共有して楽しむようになった。その際に、ただ撮影したものを投稿するのではなく、撮影映像に BGM を付与するなどの動画編集も行いうようになった。しかし動画編集では一般的に、動画に合った音楽を自分で探したり、動画の長さに合うように音楽を調整したり、といった手間とスキルが必要となる。

本研究では、動画の印象に合った楽曲を自動付与することを目標として、動画特微量からの印象推定結果に基づいた楽曲生成手法を提案している[1]。本報告では生成したシステムの改良点と、生成楽曲の評価実験について詳しく述べる。

2. 関連研究

静止画や動画に印象の合う音楽を提供する研究は旧来か

ら多く発表されている。静止画や動画、アニメーショングラフィックスを含む映像作品に BGM を付与する研究では、予め用意された楽曲の中から、映像に合った楽曲を推薦する手法[2][3][4][5]や、MIDI ファイルの音符情報を自動的に編集することで楽曲生成を行う手法がある。

MIDI 等の楽譜情報を前提として楽曲を生成する手法の例として、画像の色分布や対象物からその印象を推測し、それに合わせて楽曲をアレンジする手法[6]がある。これらと違って本研究では、ユーザの感性を学習し、ユーザの好みを反映させて楽曲を生成する。

3. 提案手法

本手法は大きく分けて 4 つの処理段階で構成される。具体的には、

(1)動画特微量: 色分布・動き分布の特微量抽出、高水準情報の抽出

(2)音楽特微量: メロディ・リズムの特微量抽出

(3)学習: 動画、メロディ・リズムの印象の関係性算出

(4)楽曲生成: ユーザの印象に合った楽曲生成

の 4 段階である。詳細について以下に論述する。

^{†1} お茶の水女子大学大学院
Ochanomizu University Graduate School
^{†2} 明治大学
Meiji University
^{†3} シドニー大学
The University of Sydney

3.1 動画特徴量

画像から受ける印象には個人差がある。文献[6]の実験結果からは、色などの低水準な特徴から印象を受ける人と、被写体などの高水準な特徴から印象を受ける人に二分されることが示唆されている。そこで、本手法では、色分布、動き分布の2種類の低水準特徴量と、被写体の種類にもとづく高水準情報を抽出し、これらの特徴量と印象との関係を学習している。

色分布、動き分布の特徴量抽出にあたって、まず動画変換ライブラリ FFmpeg (<http://www.ffmpeg.org/>) を用いて入力動画を「I フレーム」と呼ばれるキーフレームごとに分割し、分割された各動画に対して画素値およびオブティカルフローを求める。色分布および動き分布に関する特徴量抽出のためのそれ以降の処理に関しては、文献[1]の 3.1.1 項、3.1.2 項にて記載されている通りである。

高水準情報に関しては、動画から認識されると思われる被写体のうち、最も視聴者の印象に残ると思われる1つの被写体を主観的に選び、そのキーワードを手動で付与している。

3.2 音楽特徴量

本手法では、メロディとリズムを別々の素材として用意し、これらに対して文献[1]の 3.2 項に記載された音楽特徴量を算出している。

3.3 学習

続いて本手法では、動画特徴量とそれに対する各ユーザの印象の関係、またリズム・メロディの音楽特徴量とそれに対する各ユーザの印象の関係を学習する。

3.3.1 ユーザ印象評価

まずあらかじめ用意したサンプル動画、サンプルリズム・メロディを評価する際に使用する感性語対を決定する。現在採用している感性語対は文献[1]の 3.3.1 項に記載されている。そして各ユーザにサンプル動画を閲覧してもらい、またサンプルメロディ・サンプルリズムを聴取してもらい、以下の評価システム(図1参照)使用してもらうことでユーザの印象値を収集した。この評価システムに関しては文献[1]の 3.3.1 項に記載されている通りである。



図1: ユーザ評価システム

3.3.2 動画の低水準特徴量・音楽特徴量からの印象学習

3.1, 3.2 節で示した動画の低水準特徴量・音楽特徴量から印象値を推定する。本研究では文献[1]の 3.3.2 項に記載されているように、印象推定を行う際に教師なしのニューラ

ルネットワークアルゴリズムである SOM (Self Organizing Map)[7]を適用している。

3.3.3 被写体の印象値推定

高水準情報である被写体のキーワードから印象値を計算するにあたって、我々は word2vec (<https://code.google.com/archive/p/word2vec/>) を使用した。印象値算出方法は文献[1]の 3.3.3 項に記載されている通りである。

3.3.4 動画の印象値計算

3.3.2 項で求めた色分布、動き分布の印象値と、3.3.3 項で求めた被写体の印象値から、動画の印象値を求める。3.1 節の冒頭でも述べたように、画像閲覧者には低水準な特徴から動画の印象を受ける人と高水準な情報から印象を受ける人に二分されることが示唆されている。このことから、ユーザにあらかじめ、どちらの印象に左右されるかを自己回答してもらった上で、文献[1]の 3.3.4 項にある式を用いて BGM を付与したい動画の印象値を算出する。

3.4 楽曲生成

3.4.1 メロディ・リズムの合成

楽曲生成としてまず、楽曲の素材となるメロディとリズムを文献[1]の 3.4.1 項に基づき選出し、合成する。そしてメロディへ音色を付与する。音色の付与に関しては文献[1]の 3.4.2 項に記載されている通りである。以上によって生成された楽曲と動画を合成することで、動画に BGM を付与する。

4. 実験からの取り組み

我々は文献[1]の4章に記載した実験結果の考察を受けて、

- ・ リズムや動き分布に関する特徴量の再検討
- ・ 印象語との相関関係の見直し

の2点に着手した。本章ではこれを検討実験と呼ぶ。これまでの実装では先行研究で用いられてきた動画・音楽特徴量と感性語対をそのまま用いてきたが、本研究においてもこれらが有効であるか検証する必要がある。そこで我々は、感性語対を用いた印象評価結果と各特徴量との相関を求め、その結果から動画・音楽特徴量と感性語対を再選定した。具体的な選定方法は以下の通りである。

まず感性語対と動画・音楽特徴量の候補を選出した。表1に選出された感性語対と動画・音楽特徴量を示す。動画特徴量に関しては文献[1]の 3.1 節で列挙したものを採用した。音楽特徴量と感性語に関しては再度文献[8,9,10,11,12]を参考に選定した。この感性語対を用いて、図1にある評価システムを使用しユーザアンケートを実施した。アンケート回答者は情報科学を専攻する大学生・大学院生16名であった。評価システムに用いるサンプル動画・楽曲の選定には、まず動画・音楽特徴量にもとづいて K-means 法により動画・楽曲をクラスタリングし、各クラスから代表動画・楽曲を選出することで、できるだけ多彩な特徴量を有するサンプル動画群・サンプル楽曲群を選定した。この検討実

験では、50 個の動画から 10 個を選出してサンプル動画群とした。また、40 曲のメロディから 15 曲、21 曲のリズムから 10 曲を選出してサンプル楽曲群とした。

そしてアンケート回答をもとに、感性語対に関するサンプル動画・楽曲の適合度とそれらの特徴量との相関係数を算出した。その結果、表 1 にあげた感性語対、特徴量が相関の高いものとして選出された。

表 1：動画・楽曲の感性語と特徴量の候補

動画の感性語候補	動画の特徴量候補
・ 楽しい—切ない ・ 情熱的—さわやか ・ 勇ましい—かわいらしい ・ 激しい—穏やか ・ 元気—落ち着いた ・ 明るい—暗い ・ 静か—うるさい ・ 軽い—重い ・ 派手—地味 ・ 速い—遅い	・ 黒/灰色/白/茶色/赤/オレンジ/黄色/緑/水色/青/ピンク/紫 ・ 速度の平均 ・ 速度の分散 ・ 速度のヒストグラム上で度数が最大となる階級値 ・ 角度の分散 ・ 角度のヒストグラム上で度数が最大となる階級値
メロディの感性語候補	メロディの特徴量候補
・ 楽しい—切ない ・ 情熱的—さわやか ・ 勇ましい—かわいらしい ・ 激しい—穏やか ・ 元気—落ち着いた ・ 明るい—暗い	・ 音数 ・ 音域 ・ 音高平均 ・ 音高分散 ・ 音長平均 ・ 音長分散 ・ 16 分音符の割合 ・ メジャーの割合 ・ マイナーの割合
リズムの感性語候補	リズムの特徴量候補
・ 静か—うるさい ・ 軽い—重い ・ 激しい—穏やか ・ 派手—地味 ・ 速い—遅い	・ 音符数 ・ 8 分音符の割合 ・ 16 分音符の割合 ・ 3 連符の割合 ・ シンバルの割合 ・ ハイハットの割合 ・ バスドラの割合 ・ タムの割合 ・ スネアの割合

動画に関しては動きの特徴量と感性語対との相関が見られなかったことから、動きの特徴量として別の特徴量を検討した。今までは動きの特徴量として動画全体のオプティカルフローを求め、文献[1]の 3.1.2 項にあげた特徴量を抽出していた。そこで物体追跡により動いている物体のオプティカルフローを求め、同じ特徴量を抽出したところ、

感性語対との高い相関が得られたことから、この特徴量を加えて楽曲を生成するものとする。しかし風景などの動いている物体を撮影していない映像に関しては、今まで通り動画全体のオプティカルフローによって計算した特徴量を採用するものとする。最終的に決定した感性語対と特徴量を表 2 に示す。

表 2：選出された動画・楽曲の感性語と特徴量

動画の感性語	動画の特徴量
・ 楽しい—切ない ・ 激しい—穏やか ・ 元気—落ち着いた ・ 明るい—暗い ・ 軽い—重い ・ 派手—地味 ・ 速い—遅い	・ 黒 ・ 白 ・ 緑 ・ 水色 ・ 速度の平均 ・ 速度の分散 ・ 速度のヒストグラム上で度数が最大となる階級値 ・ 角度の分散 ・ 角度のヒストグラム上で度数が最大となる階級値
メロディの感性語	メロディの特徴量
・ 楽しい—切ない ・ 激しい—穏やか ・ 元気—落ち着いた ・ 明るい—暗い	・ 音数 ・ 音域 ・ 音高分散 ・ 音長平均 ・ 16 分音符の割合 ・ メジャーの割合 ・ マイナーの割合
リズムの感性語	リズムの特徴量
・ 軽い—重い ・ 激しい—穏やか ・ 派手—地味 ・ 速い—遅い	・ 音符数 ・ 8 分音符の割合 ・ 16 分音符の割合 ・ シンバルの割合 ・ バスドラの割合

5. 再実験と考察

表 2 に示した動画・楽曲の特徴量と感性語対を用いて再実験を実施し、楽曲を生成した。被験者は 4 章で述べた検討実験と同じ大学生・大学院生 16 名であった。また、4 章で示したユーザアンケート結果をそのままユーザ印象評価結果として再利用した。

本章で示す再実験では 2 種類の実験を実施した。まず本研究で使用する楽曲素材の印象値が、実際にユーザの感じる印象と一致しているかを検証した。続いて、生成した楽曲がユーザの動画への印象と合っているかを評価した。本章ではこれらの実験手順の詳細を説明し、結果と考察を述べる。

5.1 楽曲素材の印象値検証における実験結果と考察

本実験では楽曲生成に使用するメロディ 40 曲、リズム進行 21 曲のうち、メロディ 15 曲、リズム進行 10 曲を学習データ、残りのメロディおよびリズムをテストデータとする。そしてまず、学習データの音楽特徴量および印象評価結果を SOM に入力して学習させる。続いてその学習結果を用いて、テストデータの音楽特徴量を SOM に入力し、その印象値を推定する。そして推定した印象値をもとに動画の印象と近いメロディ、リズム進行を選出する。

ここで、この学習によって得られた印象値とユーザが実際に感じる印象とが一致するか、検証する必要がある。そこで被験者 16 名の各々に対して、テストデータを構成するメロディおよびリズム進行の印象値を推定する。我々はテストデータに含まれるメロディおよびリズム進行の中から 2 つずつ（以下、メロディ A、メロディ B、リズム進行 A、リズム進行 B と称する）を選出し、それらについて本手法を用いて印象値を推定した。そして、それぞれメロディ・リズム進行を実際に聞いてみて、推定された印象値と自身がもった印象が一致しているか、被験者 16 名に対して 5 段階評価を求めた。なお本実験では、すべての被験者に同じメロディ・同じリズム進行を聞かせている。ここでメロディ A、メロディ B、リズム進行 A、リズム進行 B には、特徴量間の距離が最も大きい 2 つの素材を選んだ。

まずメロディに関する実験結果を図 2、図 3 に示す。

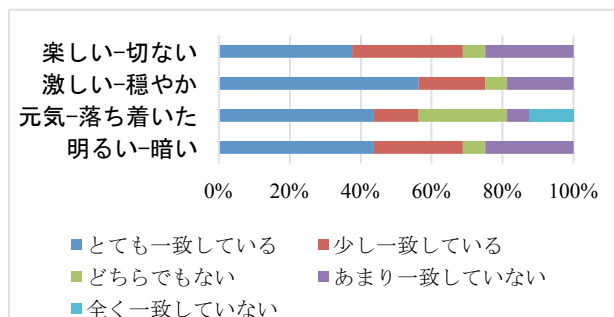


図 2：メロディ A に関する結果

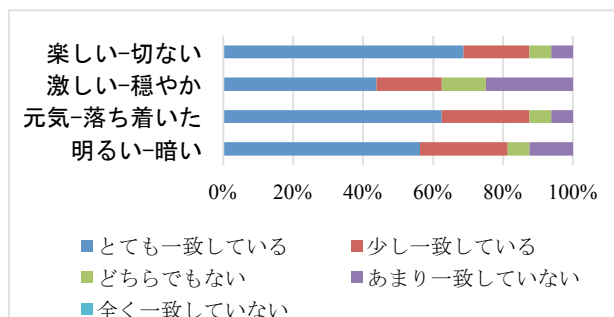


図 3：メロディ B に関する結果

メロディ A、メロディ B ともに、すべての感性語対において半数以上の被験者が「一致している」と回答した。特にメロディ B に関しては、「楽しい-切ない」「元気-落ち着いた」

いた」「明るい-暗い」の感性語対において半数以上の人が「とても一致している」と回答した。実際に被験者からは「どの感性語対も実際の印象とよくあっていた」「かなり一致している」という高評価の意見が得られた。

続いてリズム進行に関する実験結果を図 4、図 5 に示す。

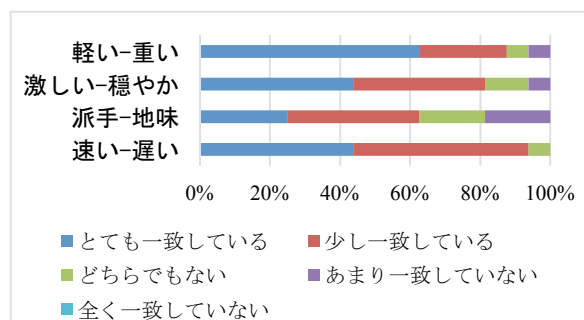


図 4：リズム進行 A に関する結果

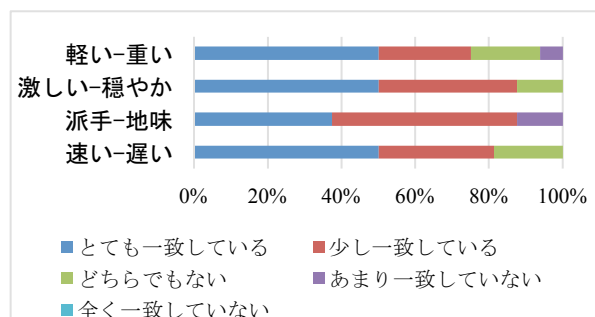


図 5：リズム進行 B に関する結果

リズム進行に関しても、リズム進行 A、リズム進行 B ともに 7 割以上の被験者が「一致している」と回答した。特にリズム進行 A の「速い-遅い」に関しては 9 割以上の被験者が「とても一致している」と回答している。

高評価の意見が多かった一方で、メロディ A の感性語対「元気-落ち着いた」に関して「全く一致していない」という評価をした被験者が 2 人(13%)いた。またメロディ・リズムともに「片方に振れ過ぎている」「そこまで一方の印象が強いわけではない」という意見もあった。一因として、類似する音楽特徴量を有するメロディやリズム進行が他にない場合や、類似する印象値を有するメロディやリズム進行が他にない場合には、SOM による印象値推定の精度が下がる傾向にあると考えられる。4 章で示した検討実験において、できるだけ多彩な特徴量を有するようにサンプル音源を選んできたが、この点についてさらに検討が必要であるように思われる。

また「本実験に用いている MIDI シンセサイザーの音質はどこか悲しいように感じられるので、『明るい-暗い』の判断をつけにくい」という意見もあった。さらに、現段階で用いているメロディ素材では音の強弱を表すベロシティがすべて一定値であるため、MIDI シンセサイザーの音質が単調な印象を与えやすいものと思われる。ベロシティに限らず MIDI 信号中の各値を調節することで、印象の異なる

る楽曲素材を多く作る必要があると考えられる。

5.2 楽曲生成に関する実験結果と考察

本実験では4章で示した検討実験に参加した16名の被験者に対し、動画の評価に関して色や動きといった低水準な特徴に印象が左右されたか、被写体のような高水準な特徴に印象が左右されたかを調査した。そして以下に示す3種類の動画に対して、

(A) 低水準な特徴の印象に重みを付けた楽曲

(B) 高水準な特徴の印象に重みを付けた楽曲

を生成し、それぞれに対し動画の印象と合っているか5段階評価で回答を求めた。そして(A)、(B)でどちらが動画の印象に合った楽曲であったか回答してもらった。被写体のキーワードを動画1では海、動画2では犬、動画3では子供として楽曲を生成した。

動画1：人がいない夕暮れの海辺の様子

動画2：犬が草むらを元気に走っている様子

動画3：公園で子供たちが遊んでいる様子

まず動画の印象が左右される方に重みを付けた楽曲の評価結果を図6に示す。

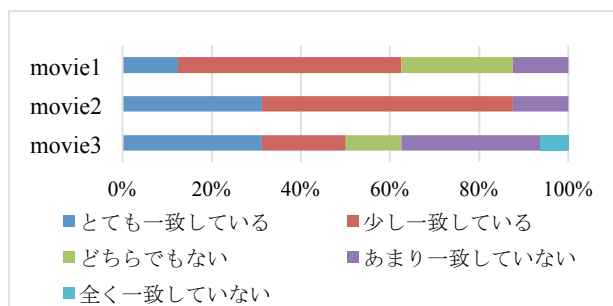


図6：印象が左右される方に重みを付けた楽曲評価結果

すべての動画において半数以上の被験者が動画の印象と生成楽曲の印象が「一致している」と回答した。動画1に対する生成楽曲の評価に関しては「犬が走っているのもう少しスピード間のある曲の方が似合いそう」といった意見があった。また動画2では「少し軽快すぎた」と感じた。動画3では「遊具のスピード感が楽曲に出れば、さらに良かったように思う」といった意見が寄せられた。このように動画の動きに関する印象が楽曲にも影響を受ける意見が多くあった。このことから、現段階で抽出しているオプティカルフローの精度向上や、動き分布と感性語対の相関を再度見直す必要がある。

続いて、動画3つのキーワードである犬、海、子供に対して3.3.3項に示した手法で印象値を推定した。出力された印象値を表3に示す。

動画1に対して(B)の高水準な印象に重みを付けた楽曲へのコメントには「元気な印象が上手く出ている」「メロディやリズムの軽快な感じがすごくあっていた」という意見があった。実際にメロディの感性語「元気—落ち着いた」、リズム進行の感性語「速い—遅い」ではどちらも「元気」「速い」に数値が振れている。一方「動画よりも重い楽曲

になったなと感じた」という意見があり、実際に「軽い—重い」に対する数値を見ると「重い」に振れている。

動画2に対して(B)の楽曲には「メロディがやや暗めなところがよかった」というコメントがあり、表の「明るい—暗い」でも「暗い」に数値が振れている。一方「落ち着いた動画に対して、少し雰囲気を楽しそうすぎる気がした」というコメントがあった。実際に「楽しい—切ない」に関しては「楽しい」に数値が振れていることからこの印象を考慮した楽曲が生成されてしまった。

動画3に対して(B)の楽曲には「ちょっとメロディと伴奏に明るさが足りない気がしました」「元気が足りないように感じた」などのコメントがあった。「元気—落ち着いた」に関しては「落ち着いた」に数値が振れており、また「明るい—暗い」は「明るい」に数値が振れられているもののその値が小さいことが原因であると考えられる。

表3：キーワードに対する印象値

感性語	犬	海	子供
楽しい—切ない	0.0634	0.0323	0.0718
軽い—重い	-0.0014	0.0021	0.01
激しい—穏やか	-0.069	-0.101	-0.03
派手—地味	-0.054	-0.015	-0.006
速い—遅い	0.0097	0.045	-0.037
元気—落ち着いた	0.044	0.048	-0.028
明るい—暗い	0.02	-0.065	0.025

これらの課題に対する改善手段として、word2vecに入力する学習文書を再検討する必要がある。また、キーワードに対してユーザが感じる印象はそれぞれ異なるものと思われる。そのため、キーワードからの印象値計算にも主観を反映させる仕組みを採用する必要があると考えられる。

次に動画の印象が左右されると回答したものは別の印象に重みを付けた楽曲の評価結果を図7に示す。

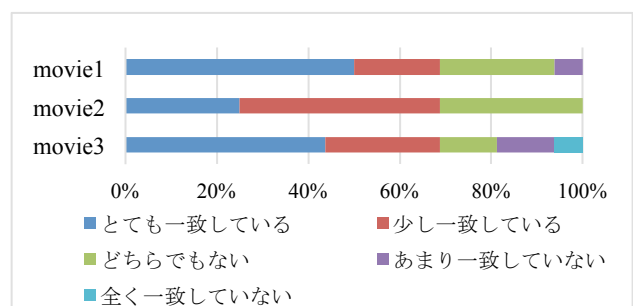


図7：印象が左右される方とは逆に重みを付けた楽曲評価結果

この結果を見ると、動画の印象が左右されると回答したものは別の印象に重みを付けた楽曲でも動画の印象と一致していると回答した被験者が半数を超えていた。さらに実際に動画の印象にあった楽曲と印象が左右される方に重

みを付けた楽曲が一致しているかを調査した結果を図8に示す。

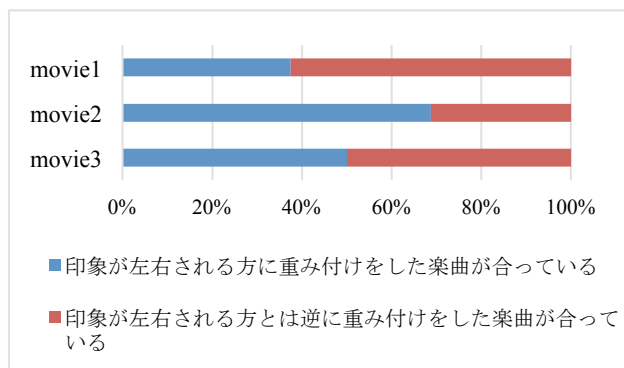


図8：生成楽曲の評価結果

動画1に関しては左右される印象とは逆に重み付けをした楽曲の方が、動画の印象とあっていると回答した被験者が半数を超えた。また動画3についても半数の被験者について同様な傾向があった。このことから、文献[1]の3.3.4項にある動画の印象値を決定する式での重み値の再検討、あるいは式そのものを再考する必要があると考えられる。

また動画を閲覧する際に動画の種類によっては低水準・高水準な特徴、どちらの印象を強く受けるか分かれる場合もあり、一概にそれらを決定するのは難しい。実際に、低水準な特徴に印象が左右されると回答した被験者のうち数名は「子供が遊んでいる様子が曲に合っている」「子供が公園で遊んでいる感じが曲から伝わる」といったように動画3のキーワードである「子供」に注目して楽曲を評価しているようだった。そのためユーザから感性データを取得する際にサンプル動画1つ1つにどちらの印象に左右されるか回答してもらうことで、この学習の結果から重み値を計算できれば、こういった課題が解決されるのではないかと考えられる。

7. まとめと今後の課題

本報告では、動画から抽出した動きや色の低水準特徴量と、被写体のキーワードである高水準情報から動画の印象を推定し、その印象評価結果に近いとされるメロディとリズムを合成することにより、動画の印象に合った楽曲を付与する手法を提案した。

6章での考察をもとに、以下を今後の課題としてあげたい。

高水準情報の自動付与。現在では動画の内容や物体を手動でタグづけしているが、一般物体認識手法の導入によりこの作業の自動化したい。その上で、一般物体認識結果にもとづく高水準情報を採用しても動画の印象を適切に推定できるかを検証したい。

高水準情報の印象値計算。現在では word2vec を使用する際に、学習データとして日本語版 Wikipedia を構成する文書群を利用しているが、よりユーザの感性にあった印象値が

出力される他の学習データを検討したい。またキーワードから受ける印象を1人1人の感性に合わせて計算する手法も検討したい。

低水準特徴と高水準情報の重みづけ。現状では文献[1]の3.3.4項の式における係数 s, t に2種類の定数値を用いており、しかも被験者の自己申告にもとづいて定数値のいずれかを採用している。被験者の入力に応じてこの値を動的に設定する方法を検討したい。

印象推定方法の精度向上。現時点での我々の実装では、限られた数のサンプル動画・サンプル楽曲への回答にもとづくユーザごとの印象推定を採用しているが、回答数が少なすぎるためにSOMによる印象推定がうまく機能していない状況が起こりえる。この問題点を解決する一手段として、サンプル動画・サンプル楽曲への回答にもとづいてユーザをクラスタリングし、クラスタごとにSOMを適用する、という方法を試みたい。

参考文献

- [1] 清水柚里奈, 菅野沙也, 伊藤貴之, 嵯峨山茂樹, 高塚正浩, “動画特徴量からの印象推定に基づく動画BGMの自動素材選出”, NICOGRAPH 2016, pp. 177-184, 2016.
- [2] C. C. S. Liem, A. Bazzica, A. Hanjalic, MuseSync: Standing on the Shoulders of Hollywood, ACM International Conference on Multimedia, pp. 1383-1384, 2012.
- [3] A. Stupal, S. Michel, Picasso-to-Sing, You Must Close Your Eyes and Draw, ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 715-724, 2011.
- [4] A. Stupal, S. Michel, Benchmarking Soundtrack Recommendation Systems with SRBench, ACM International Conference on Information and Knowledge Management, pp. 2285-2290, 2013.
- [5] J. Wang, E. Chng, C. Xu, H. Lu, Q. Tian, Generation of Personalized Music Sports Video Using Multimodal Cues, IEEE Transactions on Multimedia, Vol. 9, No. 3, pp. 576-588, 2007.
- [6] 大山喜弥, 伊藤貴之, “DIVA:画像の印象に合わせた音楽自動アレンジの一手法の提案”, 芸術科学会論文誌, Vol. 6, No. 3, pp. 126-135, 2007.
- [7] 高塚正浩, Ying Xin WU, “球面SOMのデータ構造と量子化誤差の考察およびインタラクティビティの向上”, 日本知能情報ファジィ学会誌, Vol. 19, No. 6, pp. 611-617, 2007.
- [8] 中山達喜, 吉田真一, “音楽分類における特徴量の検討”, ファジィシステムシンポジウム講演論文集, Vol. 26, pp. 1256-1261, 2010.
- [9] 菅野沙也, 伊藤貴之, “入力文書の印象と感情に基づく楽曲提供の一手法”, 情報処理学会音楽情報科学研究会, Vol. 2014-MUS-103, 2014.
- [10] 宝珍輝尚, 都司達夫, “印象に基づくマルチメディアデータの相互アクセス法”, 情報処理学会論文誌, Vol. 43 (SIG 02 (TOD 13)), pp. 69-79, 2002.
- [11] 中村均, “音楽の情動的性格の評定と音楽によって生じる情動の評定の関係”, The Japanese Journal of Psychology, Vol. 54, No. 1, pp. 54-57, 1983.
- [12] 古賀広昭, 下塩義文, 小山善文, “画像に合った音楽の選定技術”, 映像情報メディア学会技術報告, Vol. 23, No. 59, pp. 25-32, 1999.