画像メディアコンテンツの色彩情報に基づく 楽曲メディアコンテンツ印象抽出方式

木村 侑斗† 岡田 龍太郎‡ 中西 崇文†

†武蔵野大学データサイエンス学部 〒135-0063 東京都江東区有明 3-3-3
‡武蔵野大学アジア AI 研究所 〒135-0063 東京都江東区有明 3-3-3

E-mail: †yuto.kimura.2019@ds.musashino-u.ac.jp, tnakani@musashino-u.ac.jp,
‡ryotaro.okada@ds.musashino-u.ac.jp

あらまし 本稿では，画像から抽出されたその画像の印象を表す言葉(画像の印象語)と，楽曲で表そうとしているシーンや感情を表す言葉(楽曲の印象語)との相関を計量することで，画像の印象と楽曲を結び付け，画像の印象に合った楽曲とその印象を提示する方式について示す．楽曲メディアコンテンツについては，印象語とそれを表現する具体例としての楽曲とを関係づける専門家の知見がある．本方式ではそれを知識として用いることで，楽曲メディアコンテンツの特徴とその印象を紐付けする．画像メディアコンテンツについては，専門家による知見としてカラーイメージスケールを用いて，画像の色彩情報から印象語を抽出する．本方式は，入力された画像メディアコンテンツの色彩情報から，その情報に基づき印象語を抽出し，その印象語と楽曲メディアコンテンツの印象語との類似度を計算することで，画像メディアコンテンツの印象に合致した楽曲メディアコンテンツの印象を抽出すると同時にその印象を表す楽曲を提示するものである．

キーワード 異種メディア間連携, 画像メディア, 楽曲メディア, 印象語抽出

1. はじめに

近年，コンピュータの発達により扱われるメディアコンテンツの数は増大しており，多種多様な膨大なメディアコンテンツがインターネットを通じてやり取りされている．これらのメディアコンテンツに対して，例えば検索をするなど人間が操作を行う際に，これまでのような論理情報の伝達だけでは，目的を達成するのが難しくなっている．そのため，人間とコンピュータの間のインターフェイスとして，人間の要望をより正確に伝えるために，人間の感性を反映した新たなインターフェイスの実現が必要となってきている．特にこれらの多種多様なメディアコンテンツを人間が抱く感性に基づいて統一的に処理可能な機構の実現が重要となっている．これまで，文献[1]では，Kiyoki & Kitagawa によって，専門家によるメディアコンテンツの特徴から印象語とその重みで構成される印象メタデータを抽出するフレームワークである Media-lexicon Transformation Operator が提案されている．これに基づき，文献[2][3]では，にクラシック楽曲を対象としたもの，文献[4]では画像の色彩を対象としたもの，文献[5]では”音相”と呼ばれる言葉の発音情報を対象としたものが実現されている．これらの研究から，メディアデータに対応した分野の専門家の知見を導入することで，そのメディアコンテンツの特徴とその人間が喚起しうる印象語を紐付け，メディアコンテンツから印象語を抽出することを可能としている．

シーンや感情を表す言葉(印象語) と楽曲との関係を示したものとして梅垣による文献[6]がある．これは，シーンや感情などを題材として，それを表現する楽曲の具体例とともに，その印象を楽曲構造で表す際の特徴を説明している．本文献は，主にポップス音楽を作曲する際の作曲テクニックとして書かれている．文献[6]を専門家の新たな知識として，Media-lexicon Transformation Operator を適用することにより，ポップス音楽における楽曲メディアコンテンツと印象語を結びつけることが可能になると考えらえる．

本稿では，画像メディアコンテンツと楽曲メディアコンテンツの印象に基づいた連携を実現するために二つのメディアコンテンツを横断する印象抽出方式について示す．本方式は，画像メディアコンテンツから抽出された印象語と楽曲メディアコンテンツの印象を表現する印象語との相関を Word2Vec を用いて計量することで，画像の印象と楽曲を結び付け，画像メディアコンテンツの印象に合ったポップスジャンルの楽曲メディアコンテンツを提示する．本方式の画像メディアコンテンツを印象語へ変換する部分については，カラーイメージスケール[7]を用いた既存研究である文献[4]を適用する．また，楽曲メディアコンテンツと印象語を変換する部分については，梅垣による文献[6]の文献に基づくポップス音楽を対象とした新たな Media-lexicon Transformation Operator の一つとして構築する．画像からの印象の獲得に用いているカラーイメージスケールは画像の色彩情報と印象の関係を示したものであるので，構成するシステムは，入力された画像メデ

ィアコンテンツの色彩情報からその印象に合致するポップスの楽曲メディアコンテンツを提示するものとなる．

システムが提示する楽曲メディアコンテンツは梅垣による文献[6]に掲載されているものとなるが，この文献では，シーンや印象に合致した楽曲を示すと同時に，どのような特徴を持った楽曲を作ればそうした印象を表現できるのかが述べられている．こうした特徴を楽曲の特徴として抽出することができれば，楽曲メディアコンテンツ同士の類似性を計量することにより，文献で提示されている楽曲に留まらず，楽曲メディアコンテンツ一般についても検索して提示することも可能になると考えられる．さらに，著者らがこれまで取り組んできたメディアコンテンツの自動生成系である統計的一般化逆作用素の構成方式[8]と連携させることで，入力画像の印象に合致した楽曲を自動生成することが可能になると考えられる.楽曲を生成する際に，梅垣による文献で示されている特徴を楽曲構造を構成する際の制約条件として与えることが出来れば，より明確に印象を表現可能な楽曲自動生成システムが実現可能となると考えられる．本方式は，統計的一般化逆作用素の構成方式[8]によるメディアコンテンツ自動生成における“逆作用素”に対してその元となる “順作用素”にあたる，印象語抽出方式と位置付けられる．

本稿の構成は以下のとおりである．2 節では本方式の関連研究を紹介し，3 節では，梅垣による文献[7]を適用したポップス楽曲メディアコンテンツを対象とした Media-lexicon Transformation Operator 構築方式について示す．4 節では，画像メディアコンテンツの色彩情報に基づく楽曲メディアコンテンツ印象抽出方式について示す．5 節では，本方式の実装と実験を示し，6 節で本稿をまとめる．

2. 関連研究

2.1 Media-lexicon Transformation Operator

Media-lexicon Transformation Operator は，Kiyoki & Kitagawa が文献[3]により示したメディアコンテンツから印象語を抽出するフレームワークである．Media-lexicon Transformation Operator は，対象とするメディアコンテンツに関する分野の専門家による研究や評論，統計などを用いることにより，人間がそのメディアコンテンツから受ける印象語の抽出を実現する機構である．

Media-lexicon Transformation Operator は言葉同士の相関を計量する機構とセットで考案されたものである.様々な種類のメディアコンテンツを言葉という統一的なメタデータで表現し，さらにそれらのメタデータどうしの関係を計量する機構と組み合わせて，メディアコンテンツの分野をまたいだ統一的な操作を行うことを目指している．

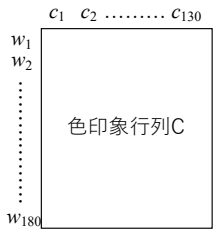


図 1 色印象行列 C

Media-lexicon Transformation Operator(ML)の一般形は次のように表される．

$$ML(Md): Md \mapsto Ws.$$

(Md: メディアコンテンツ，Ws: (重み付き)印象語群)

2.2 カラーイメージスケールとそれを用いた画像の色彩情報による Media-lexicon Transformation Operator

カラーイメージスケール[7]には，色彩と印象語の関係を表す統計データが示されている．カラーイメージスケールでは有彩色 120 色および無彩色 10 色の計 130 色の基本色に対して 180 の印象語の関連性を示している．

これまで文献[4]において，カラーイメージスケールを用いた Media-lexicon Transformation Operator が実現されている.具体的には次のような手順で構成される．

(1) 色印象行列Cの作成
基本色 130 色とそれらの印象を表す印象語 18 個の関係を表す要素からなる色印象行列Cを図 1 のように構成する．$c_1, c_2, \ldots, c_{130}$は各基本色を示し，$w_1, w_2, \ldots, w_{180}$は印象語を表す．

(2) 色彩情報の抽出
画像から色彩情報が抽出し，その色彩情報は静止画像全体における基本色 130 色の占める割合で構成される画像色彩ベクトル $\boldsymbol{m}$ によって表現される．画像色彩ベクトルを次に示す．

$$\boldsymbol{m} = (m_1, m_2, \ldots, m_{130})$$

(3) 重み付き印象語抽出
色印象行列 C，及び画像色彩ベクトル $\boldsymbol{m}$ を用いて，画像メタデータ I の抽出を行う．画像メタデータ I は，カラーイメージスケールの 180 個の印象語と同一の印象語で特徴付けられるベクトルである．

$$I = C\boldsymbol{m}$$

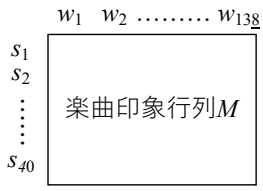


図 2 楽曲印象行列 M

3. ポップスジャンルの楽曲メディアコンテンツを対象とした印象語連携方式

本節では，梅垣による文献[7]をポップス楽曲の印象を表す専門知識と捉え，Media-lexicon Transformation Operator を応用することにより，ポップスジャンルの楽曲メディアコンテンツを対象とした印象語連携方式について示す．3.1 節では，梅垣による文献[6]の概要について述べる．3.2 節ではそれらを用いた，印象語連携実現方式について述べる．

3.1 ポップスジャンルにおける楽曲メディアコンテンツと印象語との対応

梅垣による文献[6]では，日常にある感情、風景、出来事といったイメージを系統立てて分類し，それぞれの感情、風景、出来事といったイメージごとに，どのような曲になるかの解説とともに，サンプル楽曲が楽譜として掲載されている．元々は作曲をしたい読者にどういうイメージのときにどのような曲を作曲すれば良いかを示す書籍となっている．

具体的には，本文献は大きく「恋愛イメージから作曲する」,「イベント・シーンから作曲する」,「日常シーンから作曲する」,「ゲームやアニメの世界のイメージから作曲する」,「季節のイメージから作曲する」,「グレードアップ・できるテクニックを使って作曲する」の 6 章から構成されており，最後のテクニックに関する部分を除くと全 40 シーンが挙げられている．例えば,「甘酸っぱい初恋」,「無機質にビルが立ち並ぶオフィス街」といったシーンの説明がなされている．この 40 シーンそれぞれにそのシーンに合致したサンプル楽曲が付与されている．

この文献においては，日常にある感情、風景、出来事といったイメージと楽曲が結びつけられているため，そうしたイメージを表す言葉を，楽曲の印象を表す言葉，つまり印象語と見なすことで，印象語と楽曲が関係づけられた知識として利用することが出来る．

3.2 楽曲メディアコンテンツを対象とした印象語連携実現方式

3.1 節で紹介した梅垣による文献[6]の知識を Media-lexicon Transformation Operator を応用することにより，楽曲メディアコンテンツを対象とした印象語連携方式について示す．

本方式は，日常にある感情、風景、出来事といったイメージを表す印象語を入力することで，その印象語に合致する楽曲メディアコンテンツのサンプルを導き出すものである．

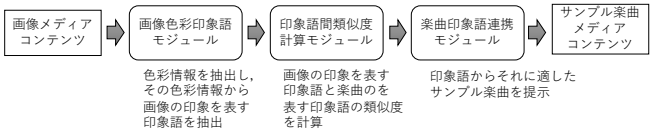


図 3 画像メディアコンテンツの色彩情報に基づく楽曲メディアコンテンツ印象抽出方式全体像

本方式は具体的には，次のプロセスで実現される．

(1) 楽曲印象行列Mの作成
文献[6]には 40 シーンが挙げられているが，これらを単語に分解すると 138 語で構成されている．これを基本印象語と定義する．基本印象語 138 語とサンプル楽曲 40 曲との関係を表す要素からなる楽曲印象行列Mを図 2 のように構成する．$w_1, w_2, \ldots, w_{138}$は印象語，$s_1, s_2, \ldots, s_{40}$は文献[6]に掲載されている楽曲サンプルを示す．

(2) 印象語ベクトル$\boldsymbol{w}$の構成
日常にある感情、風景、出来事といったイメージを表すものを重み付き印象語で表現することで，印象語ベクトル$\boldsymbol{w}$を構成する．

$$\boldsymbol{w} = (w_1, w_2, \ldots, w_{138})$$

(3) 重み付きでサンプル楽曲の相関を導出
楽曲印象行列 M，及び印象語ベクトル $\boldsymbol{w}$ を用いて，楽曲サンプル相関 R の抽出を行う．楽曲サンプル相関 R は，サンプル楽曲の相関を重み付きで表現している．

$$R = M\boldsymbol{w}$$

これによって，感情、風景、出来事といったイメージを表す印象語と楽曲サンプルとを連携することが可能である．この方式に加えて，各楽曲サンプルの特徴を抽出し，楽曲メディアコンテンツ同士の類似度を計算する方式を実現することにより，類似度による任意の楽曲メディアコンテンツを印象語ベクトルから抽出することが可能となると考えられる．

4. 画像メディアコンテンツの色彩情報に基づく楽曲メディアコンテンツ印象抽出方式

本節では，画像メディアコンテンツの色彩情報に基づく楽曲メディアコンテンツ印象抽出方式について示す．4.1 節では，本方式の概要を示し，4.2 節でそれぞれの詳細について示す．

4.1 本方式の全体像

本方式の全体像を図 3 に示す．本方式は，画像色彩印象語抽出モジュール，印象語間類似度計算モジュール，楽曲印象語連携モジュールからなる．

画像色彩印象語モジュールは，2.2 節に示すようにユーザからの画像の入力によって,色彩情報を抽出し，その色彩情報から画像の印象を表す印象語を抽出するモジュールである.印象語間類似度計算モジュールは，画像メディアコンテンツの印象を表す印象語と楽曲メディアコンテンツの感情、風景、出来事といったイメージを表す印象語という異種メディア間の印象語の類似度を計算するモジュールである．楽曲印象語連携モジュールは 3.2 節で示すように，感情、風景、出来事といったイメージを表す印象語からそれに適したサンプル楽曲を提示するモジュールである.

4.2 各モジュールの詳細

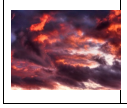
(1) 画像色彩印象語モジュール
画像色彩印象語モジュールの詳細な実現方式は，文献[4]，および 2.2 節に示している．画像から色彩情報を取り出す際，その画像メディアコンテンツ内に含まれる色がカラーイメージスケールで取り上げられている基本色 130 色のどの色と近いかの距離を求めるため，RGB 表色系から L*a*b*表色系に変換している．L*a*b*表色系上の色同士はユークリッド距離によってその色の近さを表現することができる.

(2) 印象語間類似度計算モジュール
カラーイメージスケールで定義されている画像の色彩の印象を表現する印象語と，梅垣による文献[7]による感情、風景、出来事といったイメージを表す印象語は独立して設定されており，これらのパターンマッチングで連携することは不可能である．そのため印象語間の類似度を計算するモジュールを構築することとする．具体的には，画像色彩印象語モジュールで抽出された印象語それぞれについて，3.2 節に設定した 138 語の楽曲メディアコンテンツを表現するための印象語のそれぞれの類似度を計算することとなる．本稿では，類似度計算については Wikipedia 記事を学習させた Word2vec[9]により導出することとする.

(3) 楽曲印象語連携モジュール
楽曲印象語連携モジュールの詳細な実現方式は，3.2 節に示している．これにより，感情、風景、出来事といったイメージを表す印象語に基づき，適したサンプル楽曲との相関を重み付きで表現する．相関の高いサンプル楽曲を入力した画像メディアコンテンツの印象に合致した楽曲として提示することとする.

以上のモジュールを実現することにより，画像メディアコンテンツの色彩情報に基づく楽曲メディアコンテンツ印象抽出方式が構成される．本方式により，画像メディアコンテンツの色彩に合致した印象語およびポップス楽曲メディアコンテンツを提示することが可能となる.

表 1 実験結果，入力画像と提案された印象語，サンプル楽曲の組

入力画像	最大の使用色	出力された印象語	出力された楽曲に関する語	印象語との楽曲語の距離	提案する曲
		辺鄙な	海辺	0.18597568350115756	静けさに満ちた海辺の冬
		クラシックな	オシャレ	0.17234405240195963	オシャレなカフェでくつろぐ、至福の時間
		ぜいたくな	華やか	0.17104505907961523	コスプレで華やぐ！ハロウィン
		知的な	繊細	0.18029124393475332	歴史を感じさせる重厚かつ繊細な城門
		カジュアルな	フィットネス	0.17702120857210804	黙々と汗をかくフィットネス・ジム
		精密な	繊細	0.17666705104005345	歴史を感じさせる重厚かつ繊細な城門
		緻密な	繊細	0.19703203143309625	歴史を感じさせる重厚かつ繊細な城門
		円熟した	重厚	0.1769159825758144	歴史を感じさせる重厚かつ繊細な城門
		力強い	繊細	0.17534624755850062	歴史を感じさせる重厚かつ繊細な城門

5. 実験

5.1 実験方法

4 節で示した方式を実装した実験システムを構築し，いくつかの画像メディアコンテンツを入力する．使用する入力画像は，比較的色彩がはっきりしており，かつ，単色のみで構成されていない画像とし，どのような印象語が抽出され，どのようなサンプル楽曲が提案されるかを確認した.

5.2 実験結果

具体的に入力した画像メディアコンテンツと具体的に本実験システムから出力された印象語，およびサンプル楽曲について表 1 に示す.

表 1 の 1 番目の画像は，ひまわりを中心とした画像であるが，ひまわりの花の色と周りの色が混ざってしまい，黄土色に近い色が抽出されたため，抽出された印象語も「辺鄙な」「クラシックな」といったひまわりのイメージとは逆の印象となってしまっている．そのため最終的に出力された曲も「静けさに満ちた海辺の冬」といった曲が上位に提示される．これは，色彩情報を抽出する際に，様々な色が混合してしまったことが原因でこのような結果になったと考える．これについては，2.2 節の色彩情報抽出において，主要な色を適切なクラスタ数を設定した上でクラスタリングし，それクラスタの中心色を抽出するなどの方法を適用する必要がある．これについては今後の課題とする.

表 1 の 2 番目の画像は，青い幻想的な空の画像であり，それに合致した「知的な」,「精密な」といった印象語が抽出された．これらの印象から楽曲印象として「繊細」という印象語の相関が高くなり，「歴史を感じさせる重厚かつ繊細な城門」のサンプル楽曲が提

示された．提示された曲のタイトルと画像を見比べると、「歴史」,「城門」は画像に合っているとは言い難いが、「重厚かつ繊細」は合致していると考えることができる．しかしながら一方で，色彩情報から「カジュアルな」という印象語が抽出されていることを原因として「黙々汗をかくフィットネス・ジム」という楽曲サンプルも提示されており、これはあまり画像の印象と合致していない。これは，カラーイメージスケールから抽出される印象語が本画像と合致していないことが原因であると考えられる．これについては，カラーイメージスケールから作成された色印象行列Cの学習・修正方式の実現，色彩情報以外の印象語抽出機能の実現が重要となると考えられる．これについては今後の課題とする．

表 1 の 3 番目の画像は，夕焼けに赤に染まった幻想的な空の画像である．この色彩情報から「緻密な」,「円熟した」,「力強い」といった印象語が抽出される．これらの印象語と相関のある楽曲の印象を表す印象語として,「繊細」,「重厚」といったものが抽出された．また，これらの印象語から「歴史を感じさせる重厚かつ繊細な城門」のサンプル楽曲が提示された．2 番目の画像と同様に、「歴史」,「城門」は画像の印象に合っているとは言い難いが、「重厚かつ繊細」は合致していると考えられる。

これらの結果から，画像メディアコンテンツの印象語から楽曲メディアコンテンツの印象語に連携する部分において，改善の余地はあるものの、2 番目，3 番目の結果からも印象を示すという方向性としては画像メディアコンテンツに合致した楽曲メディアコンテンツが提示できていると見ることができる．本方式を発展させていくことで、任意の画像メディアコンテンツとその印象に合致したポップス楽曲とを提示することが可能となると考えられる。

6. おわりに

本稿では，画像メディアコンテンツと楽曲メディアコンテンツの印象に基づいた連携を実現するために二つのメディアコンテンツを横断する印象抽出方式について示した．本方式によって，梅垣による文献を用いた，ポップスジャンルの楽曲メディアコンテンツを対象とした新たな Media-lexicon Transformation Operator を実現すると同時に，画像メディアコンテンツから抽出された印象語とポップスジャンルの楽曲メディアコンテンツの印象を表現する印象語との相関を計量することで，画像の印象と楽曲を結び付け，画像メディアコンテンツの印象に合ったポップスジャンルの楽曲メディアコンテンツを提示することが可能となった．

今後もインターネット上に画像メディアコンテンツやポップスジャンルの楽曲メディコンテンツが蓄積されることが予測され，これらのメディアを統合的に扱うことで，これらのメディアコンテンツの新たな利活用を促進することに可能になると考えられる．

今後の課題として，任意のポップスジャンルの楽曲メディアデータを対象とした楽曲間類似度を導入することによる印象語からの楽曲提案方式，他メディアや他の要因を対象としたメディア横断連携方式の実現，アンケート調査や大量テストデータセットによる本方式の有効性の検討が挙げられる．

参 考 文 献

[1] Kiyoki, Y. and Kitagawa, T.: Fundamental framework for mediadata retrieval system using media-lexco transformation operator, Information Modeling and Knowledge Bases, IOS Press, pp. 316-326, 2001.

[2] 吉野太智，高木秀幸，清木 康，北川高嗣:楽曲データを対象としたメタデータの自動生成とその意味的連想検 索への適用，情報処理学会研究報告，Vol.1998-DBS-116, No.2, pp.109‐116 (1998).

[3] 北川高嗣，中西崇文，清木 康:楽曲メディアデータを 対象としたメタデータ自動抽出方式の実現とその意味的 楽曲検索への適用，電子情報通信学会論文誌 D，Vol.85, No.6, pp.512‐526 (2002).

[4] 北川高嗣，中西崇文，清木 康:静止画像メディアデータ を対象としたメタデータ自動抽出方式の実現とその意味 的画像検索への適用，情報処理学会論文誌，データベー ス，Vol.43, No.12, pp.38‐51 (2002).

[5] 本間秀典，中西崇文，北川高嗣:任意の言葉を対象とした 音韻印象変換作用素の構成とその感性検索への適用，情 報処理学会論文誌，Vol.51, No.5, pp.1294‐1309 (2010).

[6] 梅垣ルナ: イメージした通りに作曲する方法 50”, 株式会社リットーミュージック(2011).

[7] 小林重順: カラーイメージスケール 改訂版, 講談社 (2001).

[8] 岡田 龍太郎, 中西 崇文, 本間 秀典, 北川 高嗣: メディアコンテンツを対象とした統計的一般化逆作用素構成 方式とその楽曲メディアコンテンツ生成への適用，情報処理学会論文誌，Vol.57, No.5, pp.1341‐1354 (2016)

[9] 鈴木正敏，松田耕史，関根聡，岡崎直観，乾健太郎: Wikipedia 記事に対する拡張固有表現ラベルの多重付与，言語処理学会第 22 回年次大会 (NLP2016), (2016).