

単語分散表現を用いた単語・フレーズ・楽曲を横断する 探索的歌詞検索システム

古屋 翔大^{†,a} 木村 和哉^{†,a} 山西 良典^{†,a}
西原 陽子^{†,a} 奥 健太^{†,b}

† 立命館大学情報理工学部 ‡ 龍谷大学理工学部

a) {is0279fk@ed, sj0035ei@ed, ryama@media, nisihara@fc}.ritsumei.ac.jp

b) okukenta@rins.ryukoku.ac.jp

概要 従来の歌詞検索システムは単語をクエリとして用いた楽曲検索であるため、ユーザの検索意図が十分に考慮されない問題がある。例えば、「恋」と入力したときに、恋という単語が含まれた楽曲を検索することは可能だが、恋を間接的に描写したフレーズは検索されない。本研究では、7万曲以上の歌詞コーパスを学習し、同一のベクトル空間で表現された歌詞の異なるレイヤーである「単語」「フレーズ」「楽曲」それぞれの分散表現を獲得する。この異なるレイヤー同士の類似度を用いることで、単語からフレーズ、フレーズから歌詞全体を検索できるなどのレイヤーを横断した探索的検索を実現する。ユーザ観察によって提案手法による歌詞検索の有用性を確認した。

キーワード 歌詞検索システム, 分散表現, 探索的検索, 評価実験

1 はじめに

楽曲の印象は、音楽のメロディーやテンポについての「音響特徴」と言語的な意味についての「歌詞印象」と大きく2つに分けることができる。音響特徴は音楽聴取における印象に対して影響を与えることは知られており [1, 2], 音響特徴を対象とした楽曲の印象分析や楽曲検索手法が数多く報告されている [3, 4, 5]。一方で、「歌詞印象」が楽曲印象において音楽聴取者に重要視されていることも明らかになっている [6]。2017年の音楽情報処理に関する国際会議 ISMIR (International Society of Music Information Retrieval) においても歌詞を題材とした研究が多く発表されている [7, 8, 9, 10]。

歌詞は図1に示すように、単語の連続によりフレーズが構成され、フレーズの連続によって楽曲が構成されている。これまでの多くの歌詞検索システムでは、単語をクエリとして用いている。しかしながら、楽曲をBGMや挿入歌として用いる場合には、楽曲全体ではなく一部のフレーズを検索したいというニーズの存在も考えられる。また、単語をクエリとして扱うのみでは、検索クエリの多義性や多様性を把握しきれずに、ユーザの検索意図が十分に考慮されない問題がある。例えば、“恋”をクエリとして入力した場合、“恋”という単語が含まれた楽曲を検索することは可能であるが、“笑顔が止まらない！ 踊るココロ止まらない！ 動き出すよ 君の元へ¹”のように恋を間接的に描写したフレーズが含まれる楽曲は検索されない。検索クエリの多義性と多様性を

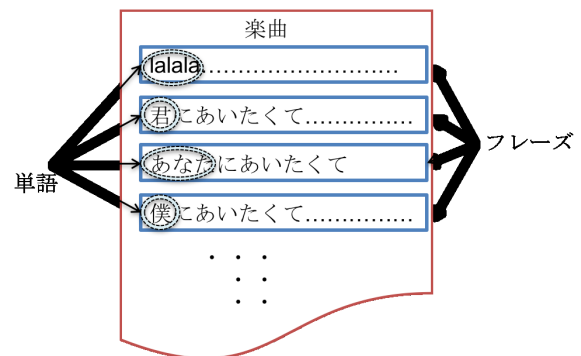


図1 歌詞の文書構造

考慮した歌詞検索を実現するためには、「短文で構成されている」「文書が特定のトピックに偏向している」「比喩的な表現が多用される」といった歌詞の文書としての特殊性を考慮した検索機構が必要であると考えられる。

本稿では、単語の分散表現を用いることで歌詞の異なるレイヤーである「単語」「フレーズ」「楽曲」を同一の単語ベクトル空間内で捉える。これにより、従来の単語からの楽曲の検索を語句一致を図る検索から意味的な類似性を捉えた検索へと発展させるのみならず、単語からフレーズ、フレーズから歌詞全体といった、歌詞のレイヤーを横断した検索の実現を図る。例えば、単語から楽曲を検索するだけでは、ユーザがクエリとして“涙”と入力したとき、「悲しい」という意味的な表現の「涙」なのか「嬉しい」という意味の涙なのかシステム側からは理解することができない。しかし、“涙”をクエリとしたフレーズの検索が実現できれば、検索結果のフレーズの中から検索者の意図に近いものを選択することで、ユー

ザが自身の検索意図に沿った楽曲に辿りつく可能性を向上させることをねらう。

2 関連研究

2.1 歌詞検索

歌詞検索に関わる既存研究として、各アーティストの固有の歌詞トピックの特徴が存在することを利用し、トピックモデル（潜在ディリクレ配分法：LDA）を用いた歌詞検索システムが挙げられる [11]。この検索システムでは、アーティストをクエリとした検索を実現しており、検索クエリに近いトピック特徴をもつアーティストを検索可能にしている。アーティストを検索対象の単位として扱っているため、少数の楽曲のみのアーティストに対してはトピックの同定が難しく検索が上手く働かない場合が考えられる。一方で、本稿では歌詞検索における多様性の確保を目指し、楽曲やフレーズを単語ベクトルとして扱い、ベクトルの類似性を利用して単語をクエリとして全ての楽曲のフレーズ検索を可能にしている。検索対象も楽曲やアーティストといった上位のレイヤーのみならず、楽曲の構成要素であるフレーズを検索可能にすることで様々な応用場面での検索における活用を図る。

歌詞検索において、楽曲中の名詞の *tf-idf* 値を利用した手法も存在する [12]。歌詞中で *tf-idf* 値の高い名詞を楽曲の特徴語として扱い、この特徴語の *tf-idf* 値を要素とするベクトルによって楽曲を表現している。このベクトルの類似性を評価することで、ある楽曲の歌詞に対して類似した歌詞が検索される。*tf-idf* 値は、一定以上の文長をもつ文書内において重要単語は繰り返し用いられる傾向にあるというアイデアに基づいた方法である [13]。しかし、歌詞では比較的短い文章であるうえに、歌詞独自の特徴である繰り返し構造（サビなど）によって一般的な文書に比べて単語の繰り返しの使用が増大する傾向にある。また、歌詞中での印象的な重要度は単語の出現頻度のみならず、組み合わせられて用いられることでの特殊性によって変化することを先行研究によって確認している [14]。そのため、*tf-idf* 値を特徴ベクトルとして用いると、頻度が低い単語に対して十分な検索ができず、歌詞集合中で一定数以上の出現頻度を持たないトピックに関連した歌詞検索性能に問題があると考えられる。

2.2 単語分散表現

Word Vector は単語の意味表現をベクトル化したものであり、単語同士の関係性をベクトル演算によって求めることができる [15, 16]。例えば、“君”に対して、“あなた”のように意味的に関連が強い単語のベクトルは、他の単語に比べてコサイン類似度が高くなる傾向にある。

Word Vector を獲得するためには、主に CBoW モデ

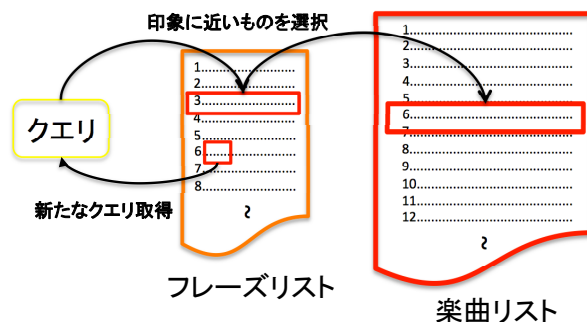


図2 探索的歌词検索のモデル

ルと Skip-gram のモデルが利用される。CBoW (Continuous Bag-of-Words) モデルでは、Word Vector 化した単語 w_t の前後に存在する $2k$ 個の単語を文脈と呼び、この文脈の Bag-of-Words 表現が入力に相当する。単語 w_t が出力層に出現する確率を求めるように、ニューラルネットワークの重みを調整しながら学習を進める。一方で、Skip-gram モデルでは、Word Vector 化した単語 w_t を入力層に与え、出力層では文脈中に出現する他の単語 $w_t + k$ を推定できるように学習を行う。本稿では、Word Vector の獲得手法において Skip-gram の手法を用いる [16]。

Word Vector は単語を対象としたベクトル化手法であるが、文章を対象としたベクトル化手法として Le によって提案された Paragraph Vector 手法がある [17]。Paragraph Vector は Word Vector を元に、文章全体を1つのベクトルで表現する。Word Vector と同様に、ベクトル化された文章は、意味的に関連が強い文章間ではベクトルが近くなる。Paragraph Vector は Word Vector を拡張した技術であるため、Paragraph Vector と Word Vector 同士のベクトル演算も可能という特徴を持つ。本稿では、Paragraph Vector を獲得するためのモデルとして PV-DBoW モデルを用いる [17]。PV-DBoW モデルの基本的な構造は Word Vector の Skip-gram と同様である。PV-DBoW モデルでは入力層に Paragraph ID を与え、出力層で文脈内の単語を推定できるように学習を繰り返し Paragraph Vector を作成する。

3 探索的な歌詞検索

歌詞検索は、「明日の天気は？」のような明確な情報要求でかつ一回の検索で十分な参照型の検索とは違い、検索者自体も検索の答えが曖昧な情報検索である。このような曖昧な検索要求をもつ歌詞検索に対して参照型方式の検索を行うと、クエリが検索者の語彙力に限られてしまうため楽曲検索の幅が狭くなり、本来の要求に合致した歌詞が検索不能となる問題がある。

そこで、図2のように、歌詞のレイヤーを超えた検索

を複数回、繰り返して行うことで検索者の検索要求自体を明確化すれば、直感的・感性的に合致した歌詞検索の実現性が高まると考えた。例えば、単語をクエリとして検索されたフレーズを閲覧することで、自身の検索要求をより明確に捉えるクエリを検索行動の中で獲得可能となる。また、単語から直接楽曲を検索するのではなく、フレーズを介して検索することで、ユーザの検索意図をより具体的に反映した検索が期待される。例えば、“恋”をクエリとして検索されたフレーズから、自分の検索時の印象に近いフレーズを選択することで、“恋”についての様々な情景の中から、ユーザの意図を反映した歌詞検索が可能になると考える。これらのことから、歌詞検索に「単語」「フレーズ」「歌詞」を横断した探索的検索 [18] を提供することで、ユーザの検索意図を考慮し、多様性を担保した感性的な歌詞検索の実現が期待される。

3.1 歌詞構造の横断的な検索

提案手法では、「単語」「フレーズ」「楽曲」の分散表現を同じベクトル空間で獲得する。これらの歌詞構造中の異なるレイヤー同士の類似度を用いることで、単語からフレーズ、フレーズから歌詞全体を検索可能にするレイヤーを横断した探索的検索を提案する。

本研究では、「うたまっぷ²」に掲載されている歌詞の中で、日本語の文字種（漢字、ひらがな、カタカナ）を含まない歌詞を除いた 75,731 曲を歌詞コーパスとして利用した。歌詞の形態素解析には MeCab [19] を、辞書には mecab-ipadic-NEologd³ を追加辞書として使用した。作成した歌詞コーパスを学習することで、同一のベクトル空間で表現された歌詞構造の異なるレイヤー「単語」「フレーズ」「楽曲」それぞれの分散表現を獲得すると同時に、「単語-フレーズ間」と「単語-楽曲間」のベクトルモデルをそれぞれ作成する。

3.2 単語からのフレーズ検索

単語からのフレーズ検索では、図 3 のように、単語をクエリとして入力することで、クエリのベクトルに対して類似度が高いベクトルをもつフレーズを出力する。例えば、クエリとして“愛”を入力すると、“愛”の単語ベクトルとのコサイン類似度が高いベクトルをもつフレーズとして“恋に溺れ それなのに”や“淡くせつない恋心 ラズベリードリーム”といったフレーズが出力される。

2 つ以上の単語を用いたフレーズ検索も可能である。例えば、クエリとして“悲しい AND 愛”と入力すると、“悲しい”と“愛”のそれぞれの単語ベクトルから算出される平均ベクトルとコサイン類似度が高いベクトルのフレーズとして“苦しい 悲しい 寂しい 怖い”や“苦

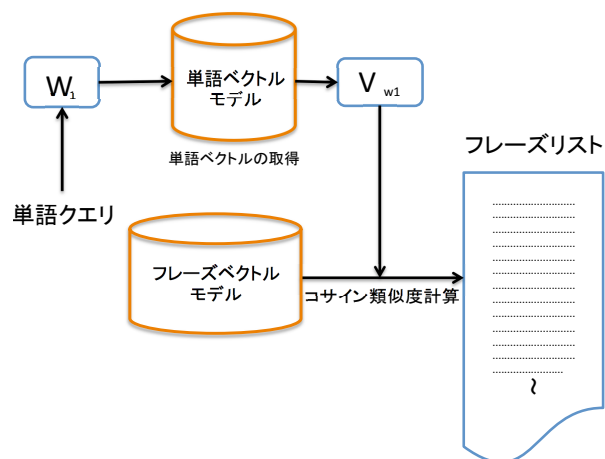


図 3 単語からのフレーズ検索

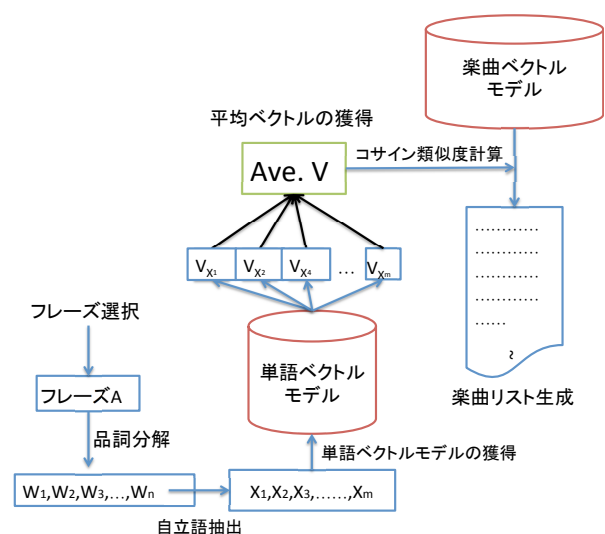


図 4 フレーズからの楽曲検索

しいのにそれなのに どうしても”などのフレーズが出力される。なお、単語からの楽曲検索においては、図 3 中で参照するベクトルモデルが楽曲のベクトルモデルとなる。

3.3 フレーズからの楽曲検索

フレーズからの楽曲の検索では、「単語-楽曲間」のベクトルモデルを応用する。フレーズからの楽曲の検索における手順を、図 4 に示す。

まず、ユーザが単語をクエリとして検索されたフレーズのリストから任意のフレーズを選択する。次に、選択されたフレーズを形態素解析して自立語を抽出する。得られた各自立語のベクトルを「単語-楽曲間」のベクトルモデルを形成している単語の分散表現から取得する。フレーズ中全ての自立語についての平均ベクトルを、選択されたフレーズのフレーズベクトルとして扱う。このフレーズベクトルに対して高い類似度を持つ楽曲ベクトルを求めることで、フレーズをクエリとした楽曲検索を実現する。

²<http://www.utamap.com/>

³<https://github.com/neologd/mecab-unidic-neologd/>

表1 実験用クエリ単語

頻出度	品詞	単語
高頻出	名詞	心, 光
	動詞	変え(る), 重ね(る)
	形容詞	切ない, 早く
	形容動詞	困難, 不幸
中頻出	副詞	かなり, ひとりで
	名詞	岸, グリーン
	動詞	育ち, 撮っ(た)
	形容詞	寒い, 濃い
	形容動詞	高級, 大キライ
低頻出	副詞	ビリビリ, ニコニコ
	名詞	サイパン, 丸印
	動詞	とびかう, 取り直し
	形容詞	懐かしく, ぼかぼかしい
	形容動詞	密やか, 怪訝
	副詞	シャンシャン, ましてや

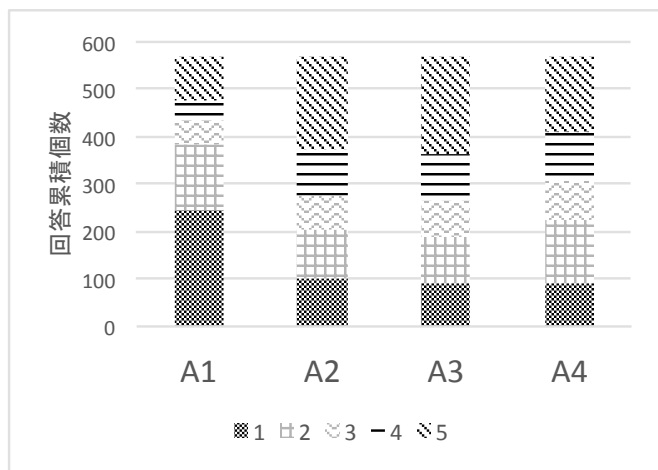


図5 単語とフレーズの印象評価実験結果

4 評価実験

従来の歌詞検索システムにはない「単語からのフレーズ検索」と「フレーズからの楽曲検索」それぞれの機能について、それぞれの実験を通して性能を確認した。

4.1 実験1：単語からのフレーズ検索

歌詞コーパス中に出現する単語のうち「名詞（固有名詞，サ変接続，一般，代名詞のみ）」「動詞（自立語のみ）」「形容詞（自立語のみ）」「形容動詞」「副詞」の単語を抽出し、各品詞での単語出現頻度で高・中・低の3グループを作成した。高頻出単語グループは歌詞コーパス中の単語出現頻度上位100個，中頻出単語グループは単語の品詞別平均出現頻度から上位100個，低頻出単語グループは歌詞コーパス中での単語出現頻度が10以上の単語のうち下位100個それぞれの単語によって構成される。ここで、中頻出単語グループの平均出現頻度は、品詞別の単語出現頻度の合計を品詞別単語数で除算したものとした。また、歌詞コーパス中での単語出現頻度が2以下の単語は、本稿での分散表現の学習時に無視される単語であるため、考慮しないものとした。各グループ

から無作為に2単語ずつ選出し、表1に示す合計30単語を実験用のクエリとして用意した。

提案システムに表1に示した各単語を検索クエリとして用いて検索されたフレーズに対して、クエリとの印象の近さを基に検索性能の妥当性を確認した。ここで、検索されたフレーズについては各単語で検索される上位200件のうち以下の条件で被験者に提示するフレーズを選出し、合計360フレーズを用意した。

A1 クエリとのベクトル類似度が最も高い3フレーズ

A2 クエリとのベクトル類似度が上位1/2に該当する3フレーズ

A3 クエリとのベクトル類似度が最も低い3フレーズ

A4 フレーズ中にクエリ単語が含まれず、クエリとのベクトル類似度が高い上位3フレーズ

被験者は、クエリとフレーズの印象の近さを5段階評価（1：かなり近い，2：近い，3：どちらとも言えない，4：遠い，5：かなり遠い）で評価した。

図5に、実験結果として、全被験者による5段階評価ごとに色分けされた累積回答数をA1 - A4のグループに分けたグラフを示す。A1の評価値1は247件，評価値2は136件で最も回答数が多く，評価値5は91件で最も回答数が少なかった。このことから検索上位にはクエリに対して印象の近いフレーズが検索されていることを確認した。A4の評価値1と2の回答数ともに92件と134件であった。これらはA1と比べると少ない。A2（評価値1：103件，評価値2：99件）やA3（評価値1：93件，評価値2：95件）に比べれば評価1は差があまりないが評価2については多く，評価値5の回答数もA2（評価値5：194件）とA3（評価値5：209件）に比べて少ない。これらのことから、クエリ単語が含まれていなくても，検索上位として出力されるフレーズは単語クエリに対して比較的，印象が近いフレーズであることが確認された。A1～A4のそれぞれのベクトル間のコサイン類似度の平均はA1は0.57，A2は0.46，A3は0.44，A4は0.52であり， $A1 > A4 > A2 > A3$ であった。これと評価実験で得られたA1～A4の被験者による印象の近さの順位が一致していることからベクトル類似度が高くなるほど，印象が近くなる傾向があると示唆された。

検索クエリの性質について考察すると，歌詞コーパス中での出現頻度が低い単語をクエリとした場合には，高頻出の単語をクエリとした場合に比べて，クエリとフレーズの印象が遠いと判断される傾向にあった。A4の結果について，被験者に印象が近いと評価されたクエリとフレーズの組み合わせの例として，“困難（な）”を

表 2 実験用クエリフレーズ

フレーズ	曲名/アーティスト
「未練 未練 未練揺さぶる 風岬」	風岬 / 神野美伽
「遠い灯台の灯(ひ)よ」	ファンレター / AKB48
「生まれ深川 住吉育ち」	三味線やくざ / 島津亜矢
「一度も一緒に写真撮ってくれなかった」	私を許さないで / 三枝夕夏 IN db
「あなた寒いわ いのちが寒い」	夕霧海峡 / 石原詢子
「汗ばむホテルルーム」	Baby Bang! feat. SPHERE / URATA NAOYA
「儲けた金で高級ゾラ買え」	まずは空手チョップ/ 餓鬼レンジャー
「アイサツせんヤツ大キライ!!!」	アイサツはハイセツよりタイセツ / グループ魂
「もうチョット ほんのチョット」	推・し・ご・と / 上原麻衣ト
「ワクワクウキウキ♪ だから私大スキ!」	変身! / 池田彩

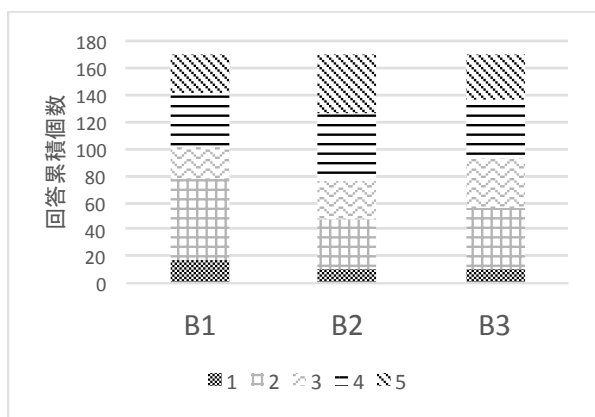


図 6 フレーズと歌詞の印象評価実験結果

クエリとした場合の「挫折こそ変われるチャンス⁴」や、「心」をクエリとした「強く惹かれる⁵」が挙げられる。これらの結果から、ユーザの検索意図に対して一定以上の性能でクエリが直接含まれるフレーズ、および、クエリが含まれないフレーズ検索が可能であることが示唆された。

4.2 実験 2：フレーズからの楽曲検索

任意の 10 フレーズをクエリとし楽曲を検索し、フレーズと検索された楽曲の印象の近さを評価した。実験に用いるフレーズを得るために、4.1 節の実験で用いた表 1 の中頻出グループの 10 単語をクエリとして用いた。各クエリに対するベクトル類似度が高いフレーズ上位 100 個のフレーズの中から無作為に 1 フレーズを選出する。表 2 に、実験に用いたクエリフレーズを示す。

⁴Dear My Life / THE ALFEE
<http://www.utamap.com/showkasi.php?url=B21680>
⁵FARAWAY / 宇徳敬子
<http://www.utamap.com/showkasi.php?url=S05566>

各クエリフレーズについて、ベクトル類似度が高い上位 200 件の楽曲から以下の条件にあてはまる楽曲を選出し、呈示楽曲とした。

B1 クエリに対して最もベクトル類似度が高い楽曲

B2 クエリとのベクトル類似度が上位 1/2 に該当する楽曲

B3 クエリとのベクトル類似度が最も低い楽曲

被験者は、クエリとして用いたフレーズと検索された楽曲の歌詞の印象の近さを 5 段階評価（1：かなり近い、2：近い、3：どちらとも言えない、4：遠い、5：かなり遠い）で評価した。

図 6 に、実験結果として、全被験者による 5 段階評価ごとに色分けされた累積回答数を B1 - B3 のグループに分けたグラフを示す。B1 は、評価値 1 は 17 件、評価値 2 は 60 件で最も回答数が多く、評価値 5 は 29 件で最も回答数が少なかった。B2 と B3 に関しては、ベクトル間の類似度が低い B3（評価値 1：10 件、評価値 2：46 件、評価値 5：33 件）のほうが B2（評価値 1：10 件、評価値 2：37 件、評価値 5：44）に比べて評価値 2 の回答数が多く、評価値 5 の回答数が少なかった。B1 について、フレーズと楽曲のベクトル間の類似度の平均が 0.51 に対し、B2 と B3 の類似度の平均はそれぞれ 0.37 と 0.34 で低い値を示し、B2 と B3 のグループ間には大きな差異が見られなかった。実験で呈示された歌詞にはクエリフレーズは含まれおらず、印象が近いと判断されている歌詞は全体としてクエリとして用いたフレーズと近似した印象をもつことが示唆された。例えば、「ワクワクウキウキ♪ だから私大スキ!⁶」をクエリとした検索結果として、全体として明るい印象を与える「ショコラティエヌ de 3 時 / 佐倉綾音⁷」が検索されたり、「遠い灯台の灯(ひ)よ⁸」をクエリとした検索した結果として悲しい印象の「遠い町 / 椎名へきる⁹」が検索されたりした。これらのことから、提案手法は従来の単語のパターンマッチングによる検索に比べて、多様でありながらも検索クエリに対して印象の近い楽曲検索を実現可能であることが示唆された。

5 おわりに

本稿では、単語からフレーズ、フレーズから楽曲全体など歌詞の構造レイヤーを横断した探索的歌词検索手法を提案した。提案手法の歌词検索は、同じベクトル空間

⁶変身! / 池田彩
http://www.utamap.com/showkasi.php?url=dk100916_38
⁷<http://www.utamap.com/showkasi.php?url=k-160831-026>
⁸ファンレター / AKB48
<http://www.utamap.com/showkasi.php?url=kt1243>
⁹<http://www.utamap.com/showkasi.php?url=T00720>

上で表現される「単語」「フレーズ」「楽曲」といった歌詞の異なるレイヤーの分散表現を獲得し、それらの類似度を用いることで実現した。

評価実験では、「単語からのフレーズ検索」と「フレーズからの楽曲検索」のそれぞれの検索機能について、主観評価実験によってクエリと検索結果との印象の近さについて性能を評価した。どちらの検索機能においても、概ね良好な検索結果が得られ、提案手法の有用性が確認された。しかしながら、特に歌詞コーパス中での出現頻度が低い単語をクエリとして用いた場合には、検索が困難になる可能性が認められるケースも存在した。分散表現の獲得の際に、不要な指示語などを取り除いてベクトルモデルを作成することで、検索性能の向上が期待される。

今後の展望として、ドライブ風景やレビュー情報などの様々なコンテキストの雰囲気合致した音楽の検索に応用していく。また、提案手法を用いた場合のユーザの検索行動についても観察し、歌詞の構造を探索することがユーザの検索意図に合致した音楽検索に対して有効であるかについても検討していく。

謝辞

本研究は一部、科学研究費若手研究 B#16K21482 および挑戦的萌芽研究#15K12151 の助成のもと行われた。また、論文中では例示のために歌詞の一部を参照させていただいた。記して謝意を表す。

参考文献

- [1] Y. Song, S. Dixon, and M. Pearce, "Evaluation of musical features for emotion classification," Proc. of ISMIR 2012, 2012.
- [2] 伊藤雄哉, 山西良典, 加藤昇平, "音楽ゆらぎ特徴を用いた楽曲印象の推定," 日本音響学会誌, pp.16-21, 2012.
- [3] 奥乃博, 北原鉄朗, 吉井和佳, "楽曲の特徴量抽出と検索技術," 電気学会誌, vol.127, no.7, pp.417-420, 2007.
- [4] 斉藤優理, 伊藤貴之, "Musicube: 特徴量空間における対話型進化計算を用いた楽曲提示インタフェース," 可視化情報学会論文集, vol.34, no.9, pp.17-27, 2014.
- [5] 豊田薫, 伊藤雄哉, 山西良典, 加藤昇平, "楽曲の音響ゆらぎに基づく感性的選曲システムの提案と選曲アルゴリズムの検討," 日本感性工学会論文誌, pp.223-231, 2012.
- [6] 森数馬, "日常の音楽聴取における歌詞の役割についての研究," 対人社会心理学研究, pp.131-137, 2010.
- [7] J. Fang, D. Grunberg, D. Litman, and Y. Wang, "Discourse analysis of lyric and lyric-based classification of music," Proc. of ISMIR 2017, pp.464-471, 2017.
- [8] K. Tsukuda, K. Ishida, and M. Goto, "Lyric jumper: A lyrics-based music exploratory web service by modeling lyrics generative process," Proc. of ISMIR 2017, pp.544-551, 2017.
- [9] K.M. Ibrahim, D. Grunberg, K. Agres, C. Gupta, and Y. Wang, "Intelligibility of sung lyrics: A pilot study," Proc. of ISMIR 2017, pp.686-693, 2017.
- [10] A. Tsaptsinos, "Lyrics-based music genre classification using a hierarchical attention network," Proc. of ISMIR 2017, pp.694-701, 2017.
- [11] 佐々木将人, 吉井和佳, 中野倫靖, 後藤正孝, 森島繁生, "Lyricradar: 歌詞の潜在的意味に基づく歌詞検索インタフェース," 情報処理学会論文誌, vol.57, no.5, pp.1365-1374, 2016.
- [12] 舟澤慎太郎, 北市健太郎, 甲藤二郎, "楽曲推薦システムのための楽曲波形と歌詞情報を考慮した類似楽曲検索に関する一検討," 情報処理学会研究報告, pp.1-5, 2008.
- [13] R. Baeza-Yates, and B. Ribeiro-Neto, Modern Information Retrieval: The Concepts and Technology behind Search (2nd Edition), Addison-Wesley Professional, 2011.
- [14] 鎌田里沙子, 山西良典, 西原陽子, 福本淳一, "表現特徴に着目した歌詞的印象的フレーズ抽出," 日本感性工学会論文誌, vol.14, no.1, pp.29-35, 2015.
- [15] T. Mikolov, W. Yih, and G. Zweig, "Linguistic regularities in continuous space word representations," The Proc. of NAACL-HLT, pp.746-751, 2013.
- [16] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," The Proc. of Workshop at ICLR, 2013.
- [17] Q. Le, and T. Mikolov, "Distributed representations of sentences and documents," Proc. of ICML 2014, pp.1188-1196, 2014.
- [18] R.W. White., and R.A. Roth., Exploratory Search : Beyond the Query-Response Paradigm, Morgan and Claypool Publishers, 2009.
- [19] T. Kudo, K. Yamamoto, and Y. Matsumoto, "Applying conditional random fields to japanese morphological analysis," The Proc. of EMNLP-2004, pp.230-237, 2004.

2017年12月26日

一部訂正

P68 の図 5 のグラフに誤りがありました。訂正後は以下のようにしております。

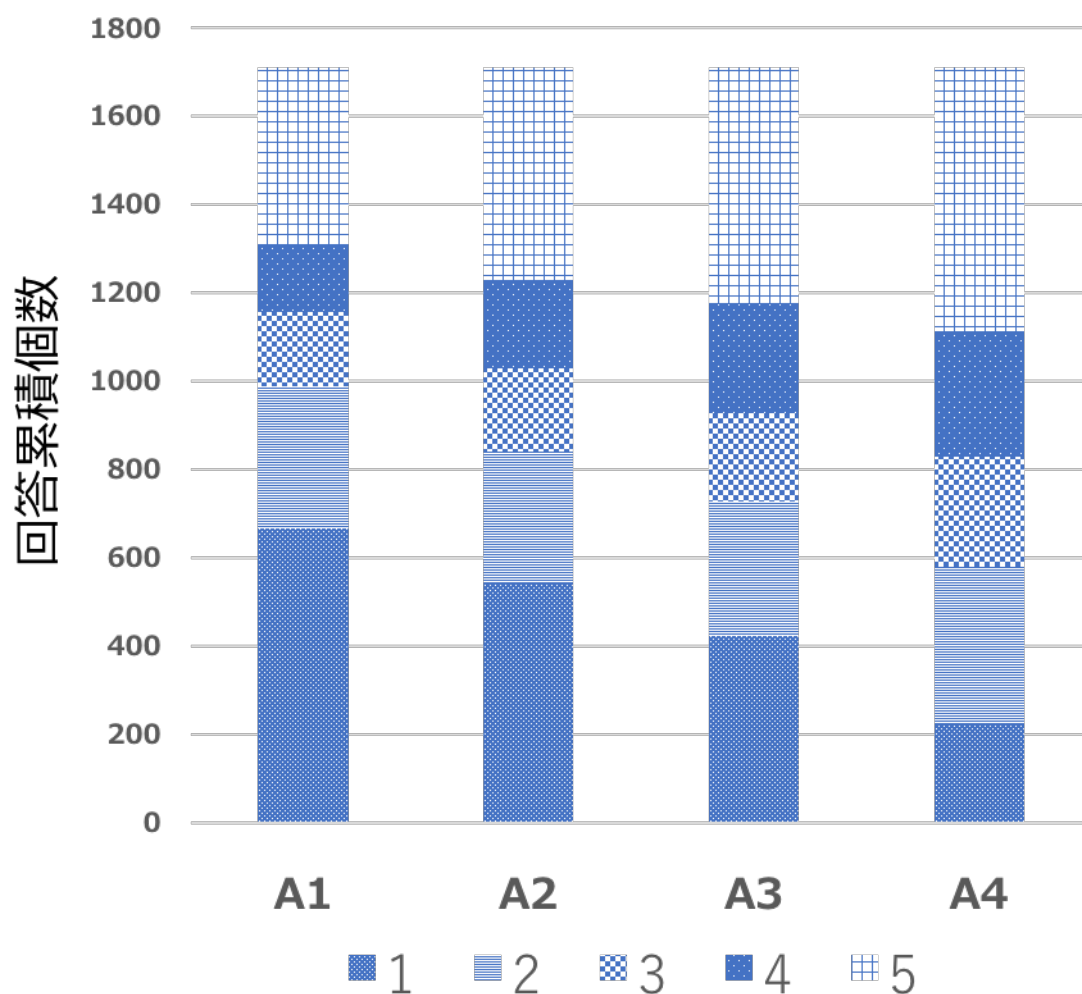


図 5 単語とフレーズの印象評価実験結果（改正後）

これに伴い、P68 ページ 4.1 節の 4 段落目の 2 文目以降の考察も以下のように変更となります。

A1 の評価値 1 は 667 件，評価値 2 は 323 件で最も回答数が多く．評価値 5 は 402 件で最も回答数が少なかった． A2 の評価値 1 は 541 件，評価値 2 は 299 件，評価値 5 は 481 件であった． A3 の評価値 1 は 424 件，評価値 2 は 304 件，評価値 5 は 532 件であった． A4 は評価値 1 は 222 件，評価値 2 は 359 件，評価値 5 は 596 件であった． このことから検索上位にはクエリに対して印象の近いフレーズが検索されていることを確認した． また，検索クエリの性質について考察すると，歌詞コーパス中での出現頻度が低い単語をクエリとした場合には，高頻出の単語をクエリとした場合に比べて，クエリとフレーズの印象が遠いと判断される傾向にあった． A4 の結果について，被験者に印象が近いと評価されたクエリとフレーズの組み合わせの例として，“困難(な)”をクエリとした場合の「挫折こそ変われるチャンス」や，“心”をクエリとした「強く惹かれる」が挙げられる． A4 は A3 と比較すると印象は遠い傾向があった． しかし，A3 と比べ全体的な傾向としては大きな差はなく，A3 は全フレーズ 1703381 件のうち類似度が高い上位 200 件付近ということから「フレーズ中にクエリ単語が含まれず，クエリとのベクトル類似度が高いフレーズ」は比較的クエリと近い印象をもつことが示唆された． これらの結果から，ユーザの検索意図に対して一定以上の性能でクエリが直接含まれるフレーズ，および，クエリが含まれないフレーズ検索が可能であることが示唆された．