

コード進行および歌詞情報を用いた 楽曲分類システムの構築

釧路高専専攻科 電子情報システム工学専攻 天元研究室 齋藤 佑樹

Construction of Musical Compositions Classification Systems Based on Chord Progressions and Lyrics

Yuki Saito

Advanced Course of Electronic and Information Systems Engineering
National Institute of Technology, Kushiro College

Abstract

In this study, we aimed to construct musical compositions classification systems with chord progressions and lyrics. We utilized Hidden Markov Model for modelizing chord progressions of musics. In order to estimate affective scores of tunes, we also applied dependency parsing to lyrics of them. We conducted the classification experiments by using data on the Web for the purpose of evaluating our systems. The results of the experiments showed that our systems classified songs in consideration of their tunes and impressions.

Keywords music information processing, machine learning, data mining, natural language processing

1. はじめに

コンピュータネットワークやデータマイニングをはじめとした情報工学の分野における技術の発展により、大量の楽曲データを取り扱うことが可能となった。また、スマートフォンや携帯音楽プレーヤーといった可搬性の高い音楽再生機器の利用により、ユーザは各自が好みとしている楽曲の視聴をより手軽に楽しむことが可能となった。これらの背景をもとに、世の中に大量に溢れている楽曲の中からユーザが求めている楽曲を検索し、取得する方法に関する研究が注目されている。

本研究に関連する研究として、楽曲におけるコード進行のパターン同士の類似度を定義し、その正当性についての検討を行なった研究 [1] や、歌詞から受ける印象に基づいた楽曲の分類および検索を行うシステムに関する検討を行なった研究 [2] などが挙げられる。楽曲を特徴付ける要因として、コード進行をはじめとした楽曲の旋律および歌詞に関する情報は重要なものとして挙げられるが、実際はこれらの要因が複雑に関係し合って 1 つの楽曲に対する印象を構成していると考えられる。また、コード進行もしくは歌詞のどちらか一方のみを反映させた楽曲分類の結果は、ユーザが希望している結果に対して齟齬が生じる可能性がある。

そこで、本研究では楽曲のコード進行と歌詞

情報の両方を利用することにより、ユーザが楽曲に対して抱く印象などをより正確に反映させることが可能となるような楽曲分類システムの構築を目標とする。また、近年では楽曲に関するさまざまな情報がインターネットを介して公開および配信されているという背景に基づき、Web 上で入手可能なデータに基づいた楽曲の分類実験を行う。そして、分類実験の結果を考察することにより本研究で構築したシステムの有効性について検証する。

本稿の構成として、まず 2 節において本研究で構築したシステムの概要について述べる。その後、3 節から 5 節にわたって本システムにおける各機能の詳細について述べる。そして、6 節において楽曲分類実験とその結果について考察し、7 節において本研究の総括を行う。

2. 本システムの概要

本研究で構築したシステムの概要および分類対象となる楽曲データについて述べる。

2.1. システム構成

本システムの構成を以下に示す。

コード進行分析部: 楽曲のコード進行に関する情報を利用し、各楽曲をコード進行に基づいてモデル化する。詳細については 3 節において述べる。

3.3. 近親調

楽曲を構成するコード進行を数値化する際に、音楽理論において定義されている近親調(関係調)[1]を利用した。近親調とは、楽曲の調同士の類似性を定めるものである。代表的な近親調の分類を以下に示す。

同主調: ハ長調(Cメジャー)とハ短調(Cマイナー)といった、同じ主音を持つ2つの調の関係を意味する。

平行調: ハ長調(Cメジャー)とイ短調(Aマイナー)といった、同一の調号を持つ2つの調の関係を意味する。

属調: ハ長調(Cメジャー)に対するト長調(Gメジャー)といった、ある調とその完全五度上にある2つの調の関係を意味する。

下屬調: ハ長調(Cメジャー)に対するヘ長調(Fメジャー)といった、ある調とその完全五度下にある2つの調の関係を意味する。

3.4. コード進行の数値化

近親調に基づいて類似している2つの調の主音は、図3.1に示す五度圏の上で近接した配置となる。例として、ハ長調(Cメジャー)を基準とした近親調について考える。平行調となるイ短調(Aマイナー)、属調となるト長調(Gメジャー)、下屬調となるヘ長調(Fメジャー)はそれぞれハ長調(Cメジャー)と隣接した位置にあるが、同主調となるハ短調(Cマイナー)は互いに直交した位置にある。本研究では、近親調に基づく主音同士の類似性と五度圏上での主音の位置関係を利用し、コード進行の数値化を行なった。例として、「G, D, C, Bm, Em, G, Am, D」というコード進行は「1, 2, 0, 2, 1, 1, 0, 2」として数値化される。

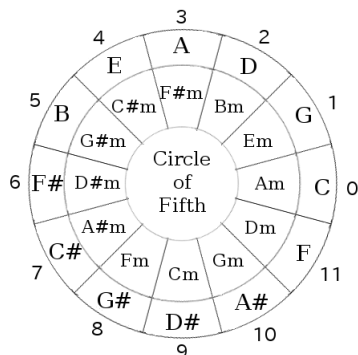


図 3.1: 五度圏

3.5. Hidden Markov Model

コード進行に基づいて楽曲をモデル化するためにHidden Markov Model (HMM)[4]を利用した。HMMは時系列で変化するデータを確率的にモデル化する手法であり、入力データの変化を上手く表現することができるパラメータ $\theta = (\pi, A, B)$ を学習により求める。ここで、 π は初期状態確率、 A は状態遷移確率、 B はシンボル出力確率を意味する。HMMのパラメータの学習法としては、入力された時系列データからEMアルゴリズムにより最適なパラメータを推定するBaum-Welchアルゴリズム[5]を利用した。

3.6. HMMによる楽曲のモデル化

HMMはモデルのトポロジーを適切に設定することで、対象とする問題に適した系列データのモデル化を行えるという特長を持つ。本研究では、楽曲のコード進行における各コードの役割がトニック、ドミナント、サブドミナントという3つに大きく分けられるという点に着目し、楽曲を図3.2に示す状態数3のErgodic HMMによりモデル化した。

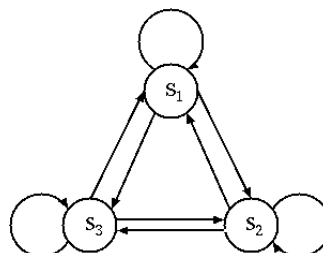


図 3.2: Ergodic HMM のトポロジー

コード進行に基づいてHMMによる楽曲のモデル化を行う際、HMMの各状態がそれぞれコードの役割(T, D, SD)に対応する。また、HMMの各パラメータはそれぞれ

初期状態確率 π : コード進行開始時のコード

状態遷移確率 A : コード遷移の様子

シンボル出力確率 B : コードの出現しやすさとして対応付けされる。

4. 歌詞情報を用いた楽曲の印象分析

本研究で構築したシステムにおける、歌詞情報を利用した楽曲の印象分析に関する手法について述べる。

4.1. Plutchik の感情の輪

印象分析を行うための感情のモデルとして Robert Plutchik により提唱された「感情の輪」[6] を用いた。感情の輪は図 4.1 に示すように、感情を 8 つの基本感情 ("ecstasy", "admiration", "terror", "amazement", "grief", "loathing", "rage", "vigilance") の強弱と組合せで表現するものである。例として, "ecstasy" の基本感情の強度を弱めたものとして "joy" と "serenity" という感情があり, "admiration" の基本感情の強度を弱めたものとして "trust" と "acceptance" という感情がある。また, それぞれの基本感情を組み合わせたものとして "love" という感情があり, これは "ecstasy" と "admiration" の基本感情が持つ性質をそれぞれ半分ずつ持つ感情として考えられる。

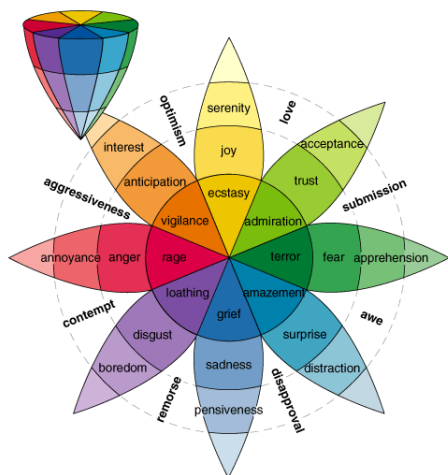


図 4.1: Plutchik の感情の輪

4.2. 感情スコアの定義

本研究では, 感情の輪に存在する英単語 (感情カテゴリ) に対し, その位置に応じた感情スコアを 8 次元の数値ベクトルとして定義した。数式で表現すると, ある感情カテゴリ c に対応する感情スコア $\text{score}(c)$ は

$$\text{score}(c) = (s_{\text{ecs}}, s_{\text{adm}}, s_{\text{ter}}, s_{\text{ama}}, s_{\text{gri}}, s_{\text{loa}}, s_{\text{rag}}, s_{\text{vig}})$$

として定義される。 $\text{score}(c)$ の各成分がそれぞれの基本感情の強度を意味する。感情の輪の内側に存在する "ecstasy", "admiration" などの感情カテゴリは, 対応する感情スコアの成分が 1.0 となり, 1 つ離れた "joy", "trust" などの感情カテゴリは, 対応する感情スコアの成分が 0.5 となる。そして, 感情の輪の最も外側に位置している "serenity", "acceptance" などの感情カテゴリ

は, 対応する感情スコアの成分が 0.25 となる。また, "love" や "submission" といった 2 つの基本感情の組合せによって構成される感情カテゴリは, 対応する感情スコアの成分の値をそれぞれ 0.5 ずつ持つものとして感情スコアを定義した。

4.3. 感情語辞書の作成

楽曲の歌詞からの印象分析を行うために, 感情に関する語句 (感情語) を格納した感情語辞書を作成した。感情語辞書は WordNet[7] および Thesaurus.com[8] から得られた単語をもとにして作成しており, 感情カテゴリと類似した意味を持つ英単語 (下位語, 類義語, 派生系など) とその和訳を含んでいる。感情語の和訳を取得する際には日本語 WordNet[9], 連想類語辞典 [10], および Weblio 類語辞典 [11] を利用した。表 4.1 に感情語辞書に格納されている感情語の一例を示す。例として, 「強い」という感情語を英訳したものは "strong" となり, これは感情カテゴリにおける "trust" というカテゴリに属する感情語となる。

表 4.1: 感情語の一例

感情語	英訳	感情カテゴリ
愛す	love	love
言葉	speech	interest
強い	strong	trust

4.4. 歌詞からの印象分析

楽曲の歌詞に対して自然言語処理の手法を適用し, その楽曲に対する印象分析を行なった。以降に処理手順の概略を示す。

4.4.1. 感情語の抽出

分類対象となる全ての楽曲の歌詞に対して係り受け解析を行い, 感情語を抽出する。係り受け解析とは, 入力として与えられた文章をその構成要素となる最小の単位 (形態素) の集合 (チャンク) に分解し, 各チャンクの修飾/非修飾の関係性を調査するための手法である。本研究では, 形態素解析を行うための解析器として MeCab[12] を, 係り受け解析を行うための解析器として CaboCha[13] をそれぞれ利用した。例として, 「愛してるの言葉で僕は強くなれる」という歌詞に対して係り受け解析を適用した結果を表 4.2 に示す。

表 4.2: 係り受け解析の適用例

ID	係り先	構成要素
0	1	愛し, てる, の
1	4	言葉, で
2	4	僕, は
3	4	強く
4	-1	なれる

表 4.2 の解析結果をもとにして感情語辞書を参照し, 感情語辞書に含まれている単語を感情語として抽出する. 上記の例では「愛す」, 「言葉」, 「強い」という感情語が得られる. なお, 係り受け解析結果の各チャンクにおける構成要素の中に「～ない」や「～せず」といった否定表現が出現する場合は, そのチャンクに含まれる感情語を抽出しないようにしている. これは, 各感情語の意味を否定させた言葉が, 必ずしもその反対の感情を持つ言葉となるとは限らないと考えられるためである.

4.4.2. 感情スコアの算出

抽出された感情語に対応する感情カテゴリを参照し, それに基づいて各感情語に割り当てられる感情スコアを算出する. 前述の例に基づいて説明すると, 表 4.1 より, 「愛す」という感情語に対する感情カテゴリとして "love" が, 「言葉」という感情語に対する感情カテゴリとして "interest" が, 「強い」という感情語に対する感情カテゴリとして "trust" がそれぞれ得られる. これらの感情カテゴリをもとに, 4.2 節において前述した感情スコアの定義に基づいて感情スコアを算出する.

4.4.3. tf-idf による感情スコアの重み付け

算出された感情スコアに対し, tf-idf による重み付けを行う. tf-idf は, 式 (4.1) に示す tf (Term Frequency) と式 (4.2) に示す idf (Inverse Document Frequency) の積により, 文書中に出現する各単語に対して重み付けを行う手法である.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (4.1)$$

$$idf_i = \log \frac{N}{N_i} \quad (4.2)$$

ここで, $n_{i,j}$ は単語 $word_i$ が文書 doc_j の中に出現する回数, N は全文書の数, N_i は単語 $word_i$ が出現する文書の数それぞれ意味す

る. tf-idf による重み付けを行うことにより, 1 つの文書に多く出現する単語に対して大きな重みを, 複数の文書に渡って多く出現する単語に対して小さな重みを割り当てることが可能となる. そのため, 1 つの楽曲の中で繰り返し出現するような感情語に対する感情スコアの重みは大きくなり, 「愛す」や「好き」などといった複数の楽曲の歌詞で多く出現することが予想される感情語に対する感情スコアの重みは小さくなる.

5. 教師なし学習による楽曲分類

本研究で構築したシステムにおける, 教師なし学習を用いた楽曲分類に関する手法について述べる.

5.1. 楽曲の特徴量

3 節で述べたコード進行の分析により得られたコード進行に関する特徴量 (HMM のパラメータ) と, 4 節で述べた歌詞情報の分析により得られた楽曲に対する感情スコア (8 次元の数値ベクトル) に基づき, 楽曲の特徴空間を構成した. 本研究では状態数 3, 出力シンボル数 12 の Ergodic HMM を利用しているため, 初期状態確率 π は 3 次元のベクトル, 状態遷移確率 A は (3×3) の行列, シンボル出力確率 B は (12×3) の行列となる. そのため, 構成される楽曲の特徴空間の次元は合計で 56 となる.

5.2. k -means++ 法

楽曲の分類を行うための手法として k -means++ 法 [14] を利用した. k -means++ 法は代表的な教師なし学習アルゴリズムの一つとして知られている k -means 法 [15] を改良した手法である. k -means 法は各データに対する初期クラスタ番号をランダムに割り当てるため, 初期クラスタの選び方によって大きく性能が変化するという欠点がある. それに対し, k -means++ 法では初期クラスタの中心を重み付けされた確率分布から選択するため, 分類開始時における初期クラスタの中心をできる限り遠くに離すことが可能となる.

6. 分類実験とその結果

本研究で構築した楽曲分類システムを評価するために行なった楽曲の分類実験およびその結果について述べる.

6.1. 実験の概要

J-Total Music における累計人気楽曲ランキングから参照した計 300 曲の楽曲データを利用し、本研究で構築したシステムによる楽曲分類実験を行なった。表 6.1 に実験対象とした楽曲データの一例を示す。なお、*k*-means++ 法により分類を行う際に指定する必要があるクラス数、本研究では五度圈上に存在する全ての主音の個数である 24 とした。

表 6.1: 累計人気楽曲ランキング (上位 5 曲)

順位	曲名	アーティスト名
1	チェリー	スピッツ
2	粉雪	レミオロメン
3	空も飛べるはず	スピッツ
4	3月9日	レミオロメン
5	ロビンソン	スピッツ

6.2. 分類実験の結果

図 6.1 および図 6.2 に分類結果として実際に構成された 2 つのクラスタを示す。以降、図 6.1 に示す結果として構成されたクラスタをクラスタ 1、図 6.2 に示す結果として構成されたクラスタをクラスタ 2 とそれぞれ称する。

Title	Artist	Original Key
花	オレンジレンジ	C
15の夜	尾崎豊	C
涙のキッス	サザンオールスターズ	C
君がいるだけで	米米CLUB	C
キズナ	オレンジレンジ	C
書く 痛く	コブクロ	C
TOO MUCH PAIN	THE BLUE HEARTS	C
フェイク	Mr.Children	Am

図 6.1: 分類結果 (クラスタ 1)

Title	Artist	Original Key
空も飛べるはず	スピッツ	C
サヨナラバス	ゆず	C
M	プリンセス・プリンセス	C
想い出がいっぱい	H2O	C
僕が僕であるために	尾崎豊	C
純恋歌	湘南乃風	C
JAM	THE YELLOW MONKEY	C
I believe	郷音	C
青春	ザ・ハイロウズ	C
Mirror	Mr.Children	C

図 6.2: 分類結果 (クラスタ 2)

次に、クラスタ 1 に含まれる楽曲に対して印象分析を行なった結果を図 6.3 に、クラスタ 2 に含まれる楽曲に対して印象分析を行なった結果を図 6.4 にそれぞれ示す。

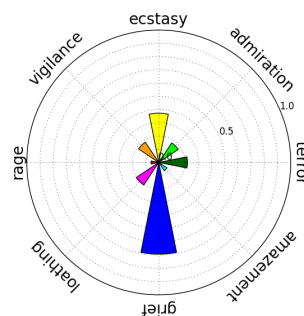


図 6.3: 印象分析結果 (クラスタ 1)

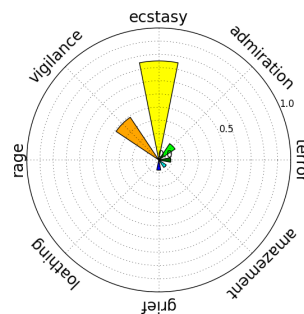


図 6.4: 印象分析結果 (クラスタ 2)

6.3. 実験結果の考察

図 6.1 および図 6.2 に示す 2 つのクラスタは、どちらもほぼ全て原曲キーを C メジャーとする楽曲の集合として構成されている。しかし、各クラスタに含まれる楽曲に対する印象分析の結果を検証したところ、クラスタ 1 は図 6.3 に示す結果のように「grief」の感情スコアが高い楽曲を多く含んでおり、クラスタ 2 は図 6.4 に示す結果のように「ecstasy」の感情スコアが高い楽曲を多く含んでいることが確認された。これらの結果より、本研究で構築したシステムは曲調の類似性だけでなく、楽曲に対する印象の類似性も反映させた分類を行なっていることが推測される。

今回の実験ではクラスタ数を 24 とした楽曲の分類を行い、その結果として各クラスタが曲調的に概ね類似している楽曲の集まりとして構成されていることが確認された。しかし、現時点では分類の際に曲調の類似性と歌詞に基づく印象の類似性のどちらを重視するかを考慮していない。そのため、今後は分類を行う際に曲調の類似性(コード進行から得られた特徴量)と印象の類似性(歌詞から得られた特徴量)に対する重み付けをユーザが指定できるようにする必要があると考えられる。また、実際にヒット曲や各ユーザが好む楽曲を分類する際には、各楽曲の曲調や印象に対して偏りが生じ

ることも予想されるため、楽曲分類を行う際の適切なクラスタ数についてもさらに検証を進める必要があると考えられる。

7. おわりに

本研究では、楽曲のコード進行に関する情報を HMM により分析し、歌詞に関する情報を用いて楽曲に対する印象を推定する機能を持つシステムを構築した。また、コード進行と歌詞情報の双方の分析結果を利用した楽曲分類実験を行い、その結果として本研究で構築したシステムが曲調および印象を考慮した楽曲の分類を行うことを確認した。

今後の展望としては、歌詞に基づく楽曲の印象分析の結果を利用した楽曲検索を行う機能などの追加により、より実用的な楽曲分類および分析を行うシステムの実現を目指す。また、実際に曲調や印象などをユーザから聴取した楽曲データを用いた分類実験を行い、本研究で構築したシステムの定量的な評価を行う。

8. 謝辞

本研究を進めるにあたり、様々な面からご指導、ご助言、ご支援を頂いた、釧路高専情報工学科所属の天元宏准教授に深く感謝し御礼申し上げます。また、研究室での日々を共に過ごした天元研究室の同輩の皆様、ならびに釧路高専専攻科での2年間を共にした皆様に厚く御礼申し上げます。

参考文献

- [1] 長澤 慎子, 渡辺 知恵美, 伊藤 貴之, 増永 良文, “近親調を用いた楽曲クラスタリングシステ

ムの構築に向けて”, 電子情報通信学会第 18 回データ工学ワークショップ論文集, 2007.

- [2] 舟澤 慎太郎, 石先 広海, 帆足 啓一郎, 滝嶋 康弘, 甲藤 二郎, “歌詞の印象に基づく楽曲検索のための楽曲自動分類に関する検討”, 情報処理学会第 71 回全国大会講演論文集, 2009.
- [3] J-Total Music, “<http://music.j-total.net/>”
- [4] Z. Ghahramani, “*An Introduction to Hidden Markov Models and Bayesian Networks*”, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 15, No. 9, pp. 9-42, 2001.
- [5] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, “*A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains*”, The Annals of Mathematical Statistics, Vol. 41, No. 1, pp. 161-171, 1970.
- [6] R. Plutchik, “*The Nature of Emotions*”, American Scientist, Vol. 89, No. 4, pp. 344-350, 2001.
- [7] WordNet, “<https://wordnet.princeton.edu/>”
- [8] Thesaurus.com, “<http://www.thesaurus.com/>”
- [9] 日本語 WordNet, “<http://nlpwww.nict.go.jp/wn-ja/>”
- [10] 連想類語辞典, “<http://renso-ruigo.com/>”
- [11] Weblio 類語辞典, “<http://thesaurus.weblio.jp/>”
- [12] T. Kudo, K. Yamamoto, and Y. Matsumoto, “*Applying Conditional Random Fields to Japanese Morphological Analysis*”, Proc. of the 2004 Conference on Empirical Methods in Natural Language Processing, pp. 230-237, 2004.
- [13] 工藤 拓, 松本 裕治, “チャンキングの段階適用による日本語係り受け解析”, 情報処理学会論文誌, Vol. 43, No. 6, 2002.
- [14] D. Arthur and S. Vassilvitskii, “*k-means++: The Advantages of Careful Seeding*”, Proc. of the eighteenth annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027-1035, 2007.
- [15] S. P. Lloyd. “*Least Squares Quantization in PCM*”, IEEE Transactions on Information Theory, Vol. 28, No. 2, pp. 129-138, 1982.