和音構成に基づく楽音の分散表現学習とその応用

筑波大学
図書館情報メディア研究科
2020年03月
森山　治紀

目次

図目次

第 1 章

はじめに

1.1 研究背景

音楽理論とは，音楽の構造や様式を統一的に説明するための知識の体系である．歴史的には，西洋音楽における作曲のための実践的規則集として発展し，体系化されてきた [1]．一方，音楽情報処理とは，工学あるいは科学の立場から音楽の構造や性質を説明することを目指す研究領域である．音楽情報処理は音楽のあらゆる側面を対象とするが，その課題は「認識」と「生成」の二つに大別できる [2]．たとえば音源分離，自動採譜，音楽情報検索などは前者，自動作曲や機械による演奏などは後者の領域といえる．これらを区別するのは，音楽を入力対象とするか出力対象とするかの違いであり，両者にまたがる研究領域もある [2]．電子デバイスやインターネットを媒体とした音楽コンテンツの利用・創作が一般化した現代では，音楽を計算機によって処理する技術や方法論の発展が社会的に求められており，音楽情報処理という研究領域の重要性が広く認められている [2]．

音楽を計算機によって処理するためには，音楽の構造やその意味を何らかの規則の元で形式的に表現しモデル化する必要がある [3]．しかし，音楽には明文化することが難しい暗黙知が多く含まれているため，その構造や性質を厳密な規則の元で形式的に記述することが難しい [3]．これは音楽情報処理の諸課題に共通して関わる問題である．

一方，自然言語処理においても類似の問題が存在する．自然言語とは，人間が互いに意思疎通を行うために用いる言語であり，自然言語を計算機によって処理する一連の技術を自然言語処理とよぶ．自然言語は単語の多義性，構文の曖昧性，文脈依存性，主観性といった性質をもつ．それゆえ，自然言語の構造やその意味について厳密な規則に基づく統一的な説明を与えることは困難であり，自然言語処理で用いられる技術や方法論はこの問題に対処することが求められる．

このように，音楽情報処理と自然言語処理には，暗黙知に起因する曖昧性や主観性といった性質が対象のモデル化を困難にするという共通の問題構造が存在する．また，音楽と自然言語はどちらも離散的かつ階層的な構造で表現されるという点でも共通している．このような類似性から，曖昧性をもつ言語の文法規則を確率的に捉える言語モデルの手法を音楽に適用する研究がこれまでに行われてきた [4]．音符や和音の系列に対するマルコフモデル [5, 6, 7] や確率的文脈自由文法 (PCFG) によるモデル [8] などはその例である．

現在の自然言語処理では，単語を固定長の実数ベクトルにより表現する分散表現 [9] を入力としたニューラルネットワークにより言語をモデル化する研究が急速に増加しており，目覚ましい成果をあげている [10]．このような背景から，音楽の構成要素である和音に対して分散表現を適用することで，曖昧性をもつ音楽の構造をモデル化する研究が行われている [11, 12, 13, 14, 15, 16]．これらの研究では，各和音のもつ機能の違いに着目しており，和音を単位とした分散表現学習を行なっている．一方，音楽理論においては音階上の各楽音の性格の違いが述べられていることから，楽音を単位とした分散表現により音楽をモデル化する手法も考えられる．しかし，これまでに楽音を単位とした分散表現学習を行う研究はほとんど行われていない．

そこで本研究では，音楽の最も基本的な構成要素である楽音に対して分散表現を適用することで，それらの音楽的関係性を捉えたベクトル空間の獲得を試みる．さらに楽音の分散表現を元に和音や楽曲といったより大きな単位の分散表現を合成し，作曲家識別や楽曲間類似度算出といった課題に活用する．その結果の妥当性について音楽学的な観点から考察を行い，楽音の分散表現を用いる手法の有用性や応用可能性について議論する．

1.2 本論文の構成

ここまで研究の背景と目的について述べた．第 2 章では，まず自然言語処理における分散表現とその文間類似度への応用に関する研究を紹介する．次に，機能和声を中心に音楽理論の基礎を概観する．そして，音楽において分散表現を適用する研究や，作曲家識別などの個別の課題に関する研究を紹介する．最後に，音楽情報処理における本研究の位置付けを述べる．

第 3 章では，楽音の分散表現学習の内容について述べる．獲得した分散表現について主成分分析などの方法で可視化し，特徴空間上での各楽音間の関係を音楽学的な観点から考察する．

第 4 章では，得られた分散表現の応用として 2 つの実験について述べる．第一に，分散表現を活用した作曲家識別について述べる．識別に用いるニューラルネットワークモデルについて詳細に述べ，得られた結果について考察を行う．第二に，分散表現に基づく楽曲間類似度について述べる．実験の内容や手法について述べ，得られた結果の妥当性について音楽学的な観点から考察を行う．

第 5 章では，第 3 章，第 4 章で得られた結果をもとに総合的なまとめを行い，音楽情報処理の諸課題における分散表現の活用可能性について議論する．その後，本研究では扱うことのできなかった課題について述べる．

第 2 章

関連研究

2.1 自然言語処理

2.1.1 シソーラスによる語彙の知識獲得

自然言語を計算機で扱うためには「単語の意味」を計算機に理解させる必要がある．計算機上で単語の意味を表現する古典的手法ではシソーラスが用いられる．シソーラスとは，同義語，類義語，上位・下位，全体・部分などの単語間の関係性を体系的に定義した辞書である．自然言語処理において最も広く利用されてきたシソーラスは WordNet[17] である．WordNet のデータベースには約 11 万 7000 の同義語グループ (synset) が含まれる．データベースは公開されており，これまでに多くの自然言語処理の研究で利用されてきた．しかし，シソーラスの活用には大きな問題点がある．それは，各単語間の関係性を全て人間が定義しラベル付けする必要がある点である．時代とともに変化する膨大な数の単語について，単語間の関係性を定義・更新していくことは困難な作業となる．また，類義語の関係にある単語間にもニュアンスの違いがある場合があり，その点まで考慮し人手で定義していくのは非常に困難である．

2.1.2 分布仮説に基づく単語の分散表現

シソーラスを用いる手法の問題点に対処するため，単語を固定長の実数ベクトルで表現する分散表現が発展した．単語をベクトルで表現する研究の多くは分布仮説 [18] に基づく．分布仮説とは，同じ文脈で出現する単語は類似の意味をもつ傾向にあるとする考え方である．この仮説に基づき，単語の共起関係を統計的に捉えることで単語間の意味的関係を的確に捉えたベクトル空間の獲得を目指す．分布仮説に基づく分散表現の手法は，カウントベースの手法と推論ベースの手法に大別される [19].

カウントベースの手法

カウントベースの手法では，コーパスとよばれる大量のテキストデータを用いる．テキスト中のある単語の周辺に出現する単語集合を「文脈」とする．ある単語とその文脈の間で各単語同士が共起する回数を配列したものを共起行列 $\boldsymbol{X}$ とよぶ．このとき $\boldsymbol{X}_{ij}$ は，単語 j のもつ文脈の中に単語 i が出現する回数を表す．スパースな共起行列 $\boldsymbol{X}$ に対して次元削減を行うこと

で各単語を表現する固定長の密なベクトルを得る．これがカウントベースの手法の基本的なアイデアである．

カウントベースの代表的な手法は Latent Semantic Indexing (LSI)[20] である．LSI では，共起行列 $\boldsymbol{X}$ について特異値分解 (Singular Value Decomposition: SVD) を用いた次元削減を行うことにより，各単語・文脈を表現するベクトルを得る．さらに確率モデルを導入することで LSI を拡張した probabilistic LSI (pLSI)[21]，Latent Dirichlet Allocation (LDA)[22] などが提案されている．

しかし，これらカウントベースの手法では計算コストの大きさが問題となる．カウントベースの手法では，コーパス全体の統計により作成された共起行列を処理することで単語の分散表現を獲得する．コーパス中で出現する語彙数は数十万を超えるのが一般的であり，それらについて巨大な共起行列を作成し処理するには多くの計算リソースが必要となる．また，新たに語彙を追加したい場合や既存の分散表現を更新したい場合にも，再度共起行列を作成し次元削減を行う一連の処理が必要となる．このようにカウントベースの手法では，計算コストの大きさや更新作業の非効率性という問題がある．これらの問題点から，カウントベースに代わる手法として推論ベースの手法が発展した．

推論ベースの手法

推論ベースの手法では，テキスト中のある位置に出現する単語をその周辺に出現する単語集合 (コンテキスト) から推論する．すなわち，コンテキストを入力として与えた時に各単語の出現確率を出力する多クラス分類問題である．何らかのモデルを用いてこの推論問題を繰り返し解くことにより，単語とコンテキストの出現パターンを学習する．大規模なコーパスを用いて，正しく推論が行われるようにモデルの学習を行い，その過程で単語の分散表現を得る．これが推論ベースの手法の基本的なアイデアである．推論問題を解くモデルにはニューラルネットワークを用いるのが一般的である．

推論ベースの手法で用いられるニューラルネットワークでは，コーパスから少量のサンプルを抜き出し，一部のデータを用いた学習を逐次的に行うのが一般的である．これにより，語彙数が大きくカウントベースの手法で処理するのが困難なケースにも対応できる．さらに推論ベースの手法では，学習済みのパラメータを初期値として再学習を行うことにより，語彙の追加や分散表現の更新を効率的に行うことが可能となる．

推論に基づく分散表現学習の最も代表的な手法は，2013 年に Mikolov らにより提案された word2vec[23] である．word2vec のアルゴリズムは 2 層のニューラルネットワークにより構成される．word2vec では 2 種類のアルゴリズムが提案されており，それぞれ continuous bag-of-words (CBOW), skip-gram とよばれる (図 2.1)．CBOW とは，テキスト中のある位置に出現する単語をコンテキストから推論するモデルである．モデルは入力としてコンテキストを受け取り，各単語の出現確率を出力する．それに対して skip-gram では，入力と出力の関係が CBOW と逆になる．すなわち skip-gram とは，テキスト中のある単語を入力とした時にその周辺に出現する単語を推論するモデルである．

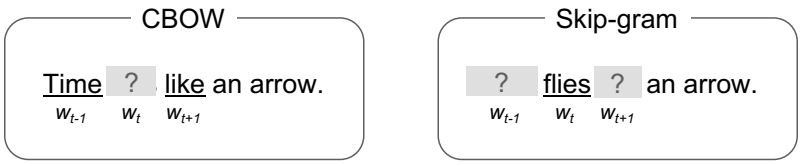


図 2.1　CBOW モデルと Skip-gram モデル

以下，モデルの詳細について説明する．テキストを単語列 $(w_1, w_2, \ldots, w_n)$ で表す．テキスト中の位置 t で出現する単語 w_t に対して，その前後 k 個の単語列をコンテキスト $C_{w_t} = (w_{t-k}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+k})$ とする．CBOW モデルでは，C_{w_t} から単語 w_t を推論する条件付き確率分布関数 $P(w_t \mid C_{w_t})$ を定義することにより，損失関数を以下の式で表すことができる．

$$L = -\frac{1}{T}\sum_{t=1}^{T} \log P(w_t \mid C_{w_t}) \tag{2.1}$$

この損失関数を最小化するようモデルの学習が行われる．大量のテキストデータを用いてこれらのモデルを学習することにより，単語の意味的関係を捉えたベクトル空間が得られる (図 2.2)．

$\overrightarrow{\text{king}} - \overrightarrow{\text{man}} + \overrightarrow{\text{woman}} \approx \overrightarrow{\text{queen}}$[24] に代表されるように，word2vec により得られる単語の分散表現は加法構成性を備えており，単語の意味的関係をベクトル演算を用いて表現できることが示されている．

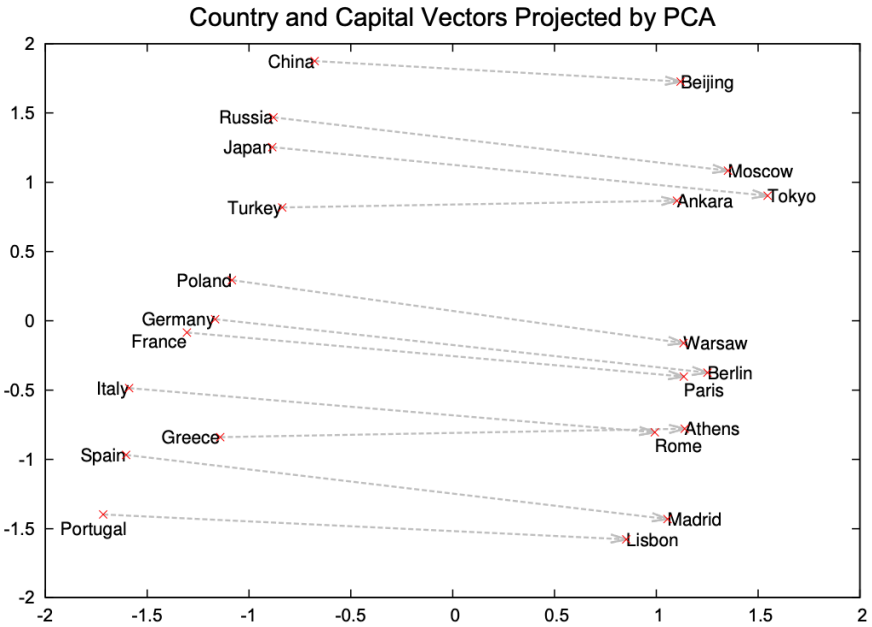


図 2.2　word2vec による分散表現 (Mikolov *et al.*, 2013[23] より引用)

Mikolov らにより word2vec が発表されて以降，推論ベースの分散表現学習手法として様々なモデルが提案されている．

GloVe[25] は Pennington らによって 2014 年に提案された分散表現学習の手法である．GloVe はカウントベースの手法で用いられる共起行列を推論ベースに組み合わせた手法である．この方法では推論による前後の文脈情報に加えて，コーパス全体の統計情報を活用することができる．

FastText[26] は Bojanowski らによって 2016 年に提案された分散表現学習の手法である．FastText では，単語をさらに細かい文字列である sub-word に分割し，これを単位として学習を行う．各単語ベクトルは，その単語を構成する sub-word ベクトルの和によって表現される．単語を単位とした手法では，コーパスに出現しない未知の単語や低頻度の単語に対応するのが困難であるのに対し，FastText ではそれらの単語についても sub-word の組み合わせにより分散表現を得ることが可能である．

これらの手法を用いて得られた分散表現の性能は，単語間の同義性や意味的関連性が事前に定義されたデータセットを用いて評価するのが一般的である [27, 28, 29]．

2.1.3 分散表現に基づく文間類似度

推論ベースの分散表現学習手法の発展に伴い，単語の分散表現に基づいて文と文の類似度を計算する手法が提案されている．

Kusner らは，word2vec を用いて得られる単語の分散表現に基づき文間類似度を計算する Word Mover's Distance (WMD)[30] を提案した．WMD では，線形計画法の輸送問題に基づいて 2 つの分布間の距離を測定する Earth Mover's Distance (EMD)[31] を用いる．

輸送問題では，供給地と需要地をそれぞれ複数設定する．それぞれの供給量もしくは需要量と，各供給地-需要地間の単位あたり輸送コストが与えられたとき，供給と需要に関する制約のもとで輸送コストの総和を最小化する輸送方法を決定する．以下に EMD の計算法を示す．

EMD では，供給地 i を特徴ベクトル p_i，その供給量を重み w_{p_i} で表し，供給地の集合を $P = \{(p_1, w_{p_1}), \ldots, (p_m, w_{p_m})\}$ とする．同様に需要地 j を特徴ベクトル q_j，その需要量を重み w_{q_j} で表し，需要地の集合を $Q = \{(q_1, w_{q_1}), \ldots, (q_n, w_{q_n})\}$ とする．供給地 i から需要地 j への単位あたり輸送コストを特徴ベクトル p_i と q_j の距離と定義し，d_{ij} で表す．供給地 i から需要地 j への輸送量を f_{ij} と定義すると，以下のような最適化問題として定式化できる．

$$\text{minimize} \quad \sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij} \tag{2.2}$$

$$\text{subject to:} \quad f_{ij} \geq 0 \qquad (1 \leq i \leq m, 1 \leq j \leq n) \tag{2.3}$$

$$\sum_{j=1}^{n} f_{ij} \leq w_{p_i} \qquad (1 \leq i \leq m) \tag{2.4}$$

$$\sum_{i=1}^{m} f_{ij} \leq w_{q_j} \qquad (1 \leq j \leq n) \tag{2.5}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \min\left(\sum_{i=1}^{m} w_{p_i}, \sum_{j=1}^{n} w_{q_j}\right) \tag{2.6}$$

この最適化問題を解くことによって得られる最適な輸送量を用いて，P, Q 間の EMD は以下の式で与えられる．

$$\mathrm{EMD}(P,Q) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}} \tag{2.7}$$

WMD は文 x から文 y への単語の輸送問題を考える．単語 i から単語 j への輸送コストを 2 つの単語ベクトル間のユークリッド距離と定義し，$c(i,j)$ で表す．文は正規化された Bag-of-Words (nBOW) ベクトル $\boldsymbol{d}$ で表現する．ここで，nBOW ベクトルの次元数は文中に出現する語彙数となり，ベクトルの各要素は対応する単語の出現頻度となる．2 つの文の nBOW 表現を $\boldsymbol{d}$, $\boldsymbol{d}'$ とし，$\boldsymbol{d}$ 内の各単語 i から $\boldsymbol{d}'$ 内の各単語 j への輸送量を配列した行列を $\boldsymbol{T}_{ij}$ とする．語彙数を n とすると，WMD の最適化問題は以下の式で定義される．

$$\min_{\boldsymbol{T}_{ij}\geq 0} \sum_{i,j=1}^{n} \boldsymbol{T}_{ij} c\,(i,j) \tag{2.8}$$

$$\text{subject to:} \qquad \sum_{j=1}^{n} \boldsymbol{T}_{ij} = \boldsymbol{d}_i \qquad (1 \leq i \leq n) \tag{2.9}$$

$$\sum_{i=1}^{n} \boldsymbol{T}_{ij} = \boldsymbol{d}'_j \qquad (1 \leq j \leq n) \tag{2.10}$$

この最適化問題を解くことによって，単語間の意味的類似性を反映した文間の類似度を計算することが可能となる．

2.2 音楽理論

2.2.1 音階

調性音楽において，起点とする音から 1 オクターブ上の音に到達するまで特定の音程関係によって配列された音列を音階という．全音階とは，5 つの全音音程と 2 つの半音音程を含む 7 つの音から構成される音階である [1]．全音階は全音音程と半音音程の順序の違いによって長音階と短音階に区別される[*1]．調性音楽では，長音階を用いるものを長調，短音階を用いるものを短調とよぶ．

調性音楽において，音階の起点となる音を主音，主音から数えて 4 番目の音を下属音，5 番目の音を属音，7 番目の音を導音とよぶ (図 2.3)．8 番目の音はオクターブ上の音にあたり，再び主音となる．これらの音は音楽的にそれぞれ異なる性格をもつとされる．主音は音階の起点となる最も重要な音であり，属音の存在によりその性格を規定される．下属音は主音と属音のはたらきを補助する役割をもつ．導音は主音への推進力を生むはたらきをする [32, 33]．このように音楽理論では，音階を構成する各楽音はそれぞれが音楽的に異なる性格をもつとされる [32]．

[*1] 全音階には長音階・短音階の他にも教会旋法などがあるが，現在の調性音楽においては長音階もしくは短音階を用いるのが一般的である．

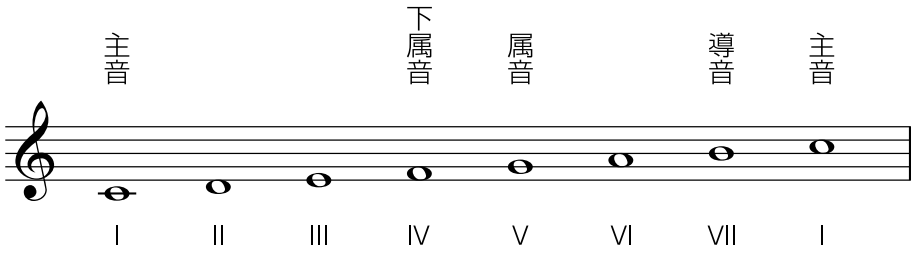


図 2.3　C を主音とした長音階

2.2.2　和音

和音とは，高さの異なる二つ以上の音が同時に響くことである．音の組み合わせによって多種多様な和音を構成することができるが，調性音楽においては三和音を基礎とする和音の考え方が最も中心的である．三和音とは，3 つの音を 3 度の音程間隔で積み重ねた和音である．特に音階の主音の上に構成される三和音を主和音，下属音の上に構成される三和音を下属和音，属音の上に構成される和音を属和音とよぶ (図 2.4)．

3 度間隔の音の積み重ねを基本として，4 つの音からなる四和音，5 つの音からなる五和音，6 つの音をからなる十一の和音，7 つの音からなる十三の和音がある．全ての和音は原則的には以上のいずれかに分類され，構成音の省略，重複，付加，変化などにより形成されると考える [33]．

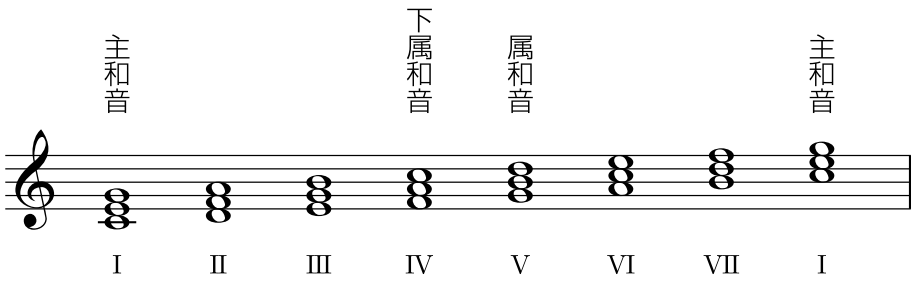


図 2.4　C を主音とした長音階上に構成される三和音

2.2.3　機能和声

和音の時間方向の配列を和声 (harmony) とよぶ．和声のパターンにより，聴く人は安定・不安定，緊張・弛緩，終止感などを感じる．調性音楽において，一般的な楽曲にはその和声を支配する基本的な原理が存在する．安定状態から不安定状態へと遷移し，その後再び安定状態へと戻るという構成である．これは主和音から始まり，不安定和音を経由して再び主和音に戻るという和声を意味する [1, 34]．このような構成をもつ和声の単位をカデンツとよぶ [34]．機能和声とは，カデンツ内の個々の和音は特定の機能をもっており，カデンツにおける和音機能の配列パターンは常に一定の型をもつとする考え方である [35]．和音の機能は Tonic (T)，Dominant (D)，Subdominant (S) の 3 種に分類され，カデンツの形成における役割の違いを

表す[*2].

Tonic
: 主音の性格を帯び，安定感を強く示す機能である．調性音楽において，この機能を最も強くもつ音階上の和音はIの和音 (主和音) である．次いでVIの和音がこの機能をもち，IIIの和音は状態によってはこの機能を果たす場合がある [32].

Dominant
: 主和音への推進力を強く示す機能である．この機能を強くもつ音階上の和音はVの和音 (属和音) である．次いでVIIの和音がこの機能をもち，IIIの和音は状態によってはこの機能を果たす場合がある [32].

Subdominant
: Tonic, Dominant とは異なる機能をもつ．この機能を強くもつ音階上の和音はIVの和音 (下属和音) である．次いでIIの和音がこの機能をもつ．Tonic や Dominant のような強い性格はないが，叙情感や開放感など状態によってさまざまな印象を与える [32].

カデンツはこれら3つの機能に基づき, 第1型: T⟶D⟶T, 第2型: T⟶S⟶D⟶T, 第3型： T⟶S⟶T のいずれかに分類される．

2.3 音楽情報処理

2.3.1 和音の分散表現

自然言語と音楽の構造的類似性に着目し，推論ベースの分散表現学習手法を音楽の構成要素に適用する研究がある [11, 12, 13, 14, 15, 16]．これらの研究では，楽曲中のある和音を入力として，その周辺に出現する和音を推論するニューラルネットワークモデルを学習することで，和音の分散表現を得ている．こうして得られた分散表現を2次元に圧縮した際に五度圏が形成されるなど，和音間の音楽的関係が反映されていることを示唆する結果 (図 2.5) も得られている [12].

これらの研究では和音を単位とした分散表現学習を行なっている．一方，音楽理論において音階上の各楽音の性格の違いが議論されるように，楽音を単位とした分散表現により音楽をモデル化する手法も考えられる．しかし，これまでに楽音を単位とした分散表現学習を行う研究はほとんど行われていない．また，和音を単位とした分散表現学習では，コーパス中に出現しない未知の和音や低頻度の和音に対応できないという課題もある．

また，単語はそれが表す意味が明確に定義されていることから，意味関係を事前に定義したデータセットを用いて分散表現を評価することができる．しかし，音楽の構成要素はそれが表す意味が明確に定義されているわけではないため，得られた分散表現を評価することが難しい点も課題である．

[*2] なお，主音および主和音はどちらも Tonic とよばれ，区別しないことが多い．これは主和音には主音の性質が強く反映されており，両者が不可分であるためである．Dominant，Subdominant についても同様である．

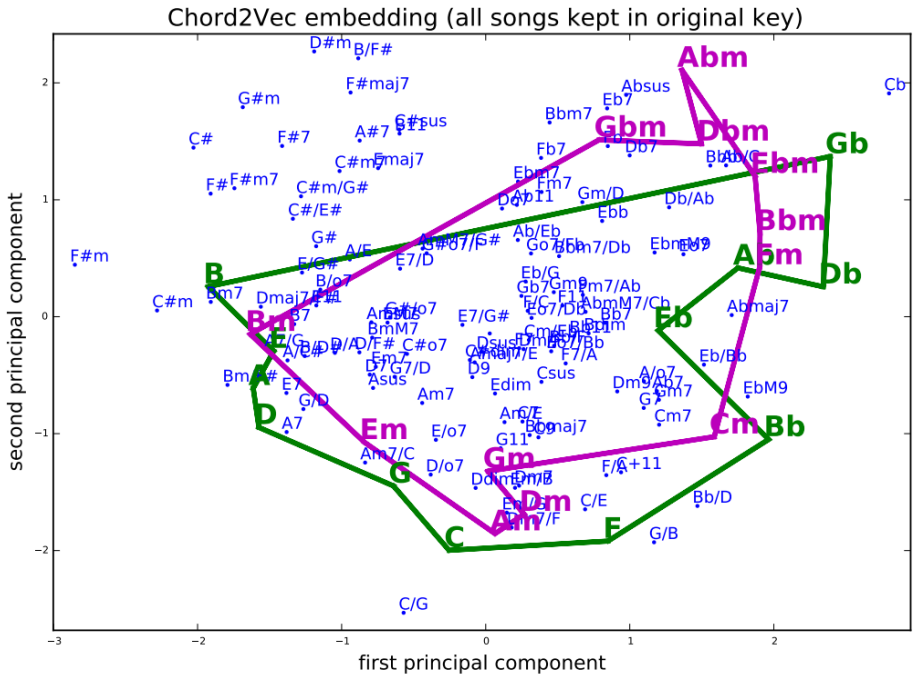


図 2.5　PCA により可視化した和音の分散表現 (Huang *et al.*, 2016[12] より引用)

2.3.2　作曲家識別

本研究に関連する研究として，楽譜情報に基づく作曲家識別がある．Pollastri らは，メロディを入力とし，Hidden Markov Model (HMM) を用いて作曲家を識別する研究を行った [36]．旋律音程と音長比を特徴量として，Mozart/Beethoven/Dvořák/Stravinsky と The Beatles の 5 クラス分類を行い，42% の識別精度を得ている．

長谷川らはラルーらの様式分析手法 [37] の定量化を試みる研究を行った [38]．様々な時代・地域のクラシック音楽を対象に，独自に定義した 14 種の特徴量を用いて 26 作曲家の正準判別分析を行い，56% の精度を得ている．判別分析の結果に対して階層的クラスタ分析を行い，共通する時代や文化をもつと考えられる作曲家が特徴空間上で近い位置に配置されることを示した (図 2.6)．

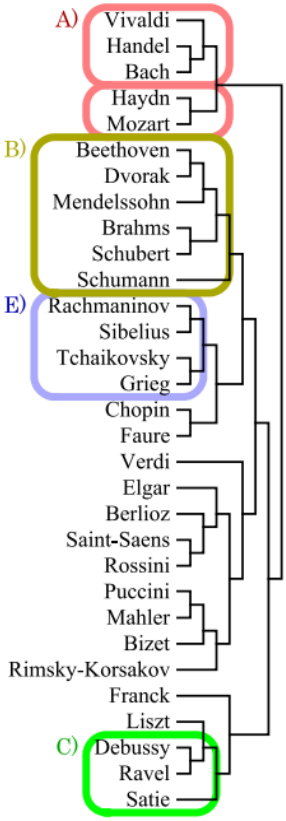


図 2.6　階層的クラスタ分析の結果 (長谷川隆ほか, 2012[38] より引用)

2.4　本研究の位置付け

音楽情報処理と自然言語処理には，暗黙知に起因する曖昧性や主観性といった性質が対象のモデル化を困難にするという共通の問題構造が存在する．また，音楽と自然言語はどちらも離散的かつ階層的な構造で表現されるという点でも共通している．このような類似性から，マルコフモデルや確率的文脈自由文法 (PCFG) などの言語モデルを音楽の構成要素に適用することで，曖昧性をもつ音楽の構造をモデル化する研究がこれまでに行われてきた [4]．近年では，単語の分散表現 [9] を用いた言語のモデル化が大きな成果をあげていることから，音楽の構成要素である和音に対して分散表現を適用する研究が行われている [11, 12, 13, 14, 15, 16]．

一方，音楽理論においては音階上の各楽音の性格の違いが議論されることから，楽音を単位とした分散表現により音楽をモデル化する手法も考えられる．しかし，これまでに楽音を単位とした分散表現学習を行う研究はほとんど行われていない．

本研究では，音楽の最も基本的な構成要素である楽音に対して分散表現を適用することで，それらの音楽的関係性を捉えたベクトル空間の獲得を試みる．さらに楽音の分散表現を元に和音や楽曲といったより大きな単位の分散表現を合成し，作曲家識別や楽曲間類似度算出といった課題に活用する．その結果の妥当性について音楽学的な観点から考察を行い，楽音の分散表現を用いる手法の有用性や応用可能性について議論する．

第 3 章

分散表現学習

3.1 概要

本章では，調性音楽で用いられる各楽音について，それらの音楽的関係性を捉えた分散表現を学習する実験について述べる．本実験では，ある時間区間で同時に鳴る楽音の集合を和音，和音の時系列を楽曲と定義する．楽曲中の和音交替のタイミングは，一つ以上の和音構成音が変化した時点とする．図 3.1 に実例を示す．図中の青い四角は楽譜上の対応する時点の和音を表しており，四角の内部は和音を構成する楽音を表す．最初の和音は C4, G4, E5 の 3 音である．2 番目の和音では，最初の和音の C4, E5 がそれぞれ D4, F5 に変化し，G4 は保持されている[*1]．以下も同様である．

これにより，離散的な記号列として楽曲を取り扱うことができるようになり，自然言語処理の手法を適用することが可能となる．本実験では，自然言語処理における分散表現学習の代表的手法である word2vec[23] を応用し，楽音の分散表現を獲得することを目指す．

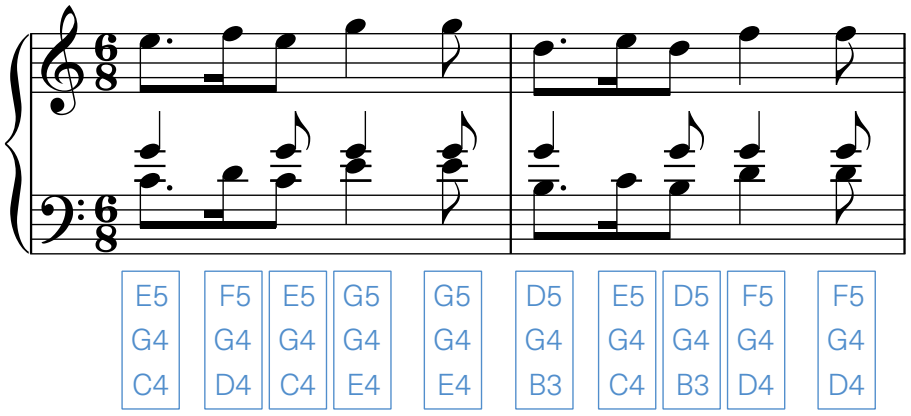


図 3.1　楽曲データから和音系列への変換

[*1] 音名は C, D, E, F, G, A, B で表記し，音名の後の数字は何オクターブ目かを表す．

3.2 手法

本実験の前提として，自然言語処理の分布仮説における「単語」と「文脈」の関係性が，音楽における「楽音」と「和音構成」の間にも成り立つと仮定する．ここで和音構成とは，和音を構成する楽音の集合を意味する．この仮定は，同じ和音構成で出現する楽音は類似の音楽的性格をもつ傾向にあることを意味する．この仮定により，分布仮説に基づく分散表現学習手法を楽音に適用し，各楽音の音楽的関係性を捉えた分散表現の獲得が可能になると期待される．

本実験では，word2vec の skip-gram モデルを応用し，楽音の分散表現学習を行う．和音の構成音の一つを入力とし，その和音を構成する他の楽音を推論するニューラルネットワークモデルを構築する (図 3.2)．このモデルを用いて，楽曲中に出現する各和音の構成音それぞれについて学習を行い分散表現を獲得する．

本実験で学習を行う楽音の分散表現は 8 次元とした．自然言語処理では，語彙数が数百万であるのに対して，各単語の分散表現は数百次元とするのが一般的である．これは低次元で単語の意味を的確に捉えた頑健性のあるベクトル表現を得るためである．本実験で取り扱う楽音の種類はピアノ音域内の 88 音であり，自然言語の語彙数と比べて数が少ない．したがって，本実験では自然言語処理で一般的な値より低い次元数を用いる．

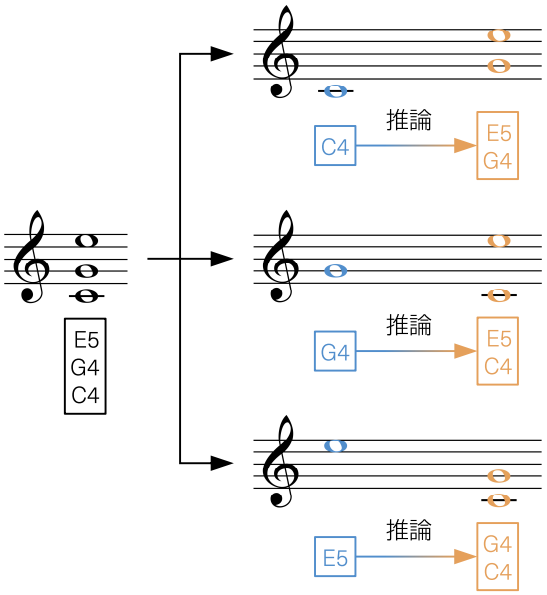


図 3.2　和音構成音を推論する skip-gram モデル

3.3 使用データ

分散表現学習には，クラシック音楽の MIDI ファイルを提供する Web サイトである Kunst der Fuge[39] よりダウンロードしたデータを使用する．Kunst der Fuge には約 19,500 の MIDI ファイルが含まれており，それらは幅広い時代・地域の作曲家について，管弦楽曲，器

楽曲，声楽曲等々様々な種類の楽曲を網羅している．各 MIDI ファイルは約 200 人の作成者のいずれかによって作成されたものである．本実験では，Kunst der Fuge に含まれる幅広い作曲家の MIDI ファイル約 8,900 曲分を学習データに用いる．

また本実験では，使用する全ての楽曲について，長調の曲はハ長調，短調の曲はイ短調に移調する．これは調性の違いが各和音の機能に与える影響を排除するためである．機能和声において，和音の機能は調の中心となる主音との相対的な関係に従って定まる．例えば，C, E, G の 3 音からなる三和音は，ハ長調では主和音 (Tonic)，ト長調では下属和音 (Subdominant)，ヘ長調では属和音 (Dominant) となる．このように調が異なれば同じ構成の和音が異なる機能を示すことになる．逆に，同一の調であれば和音構成と和音機能が 1 対 1 に対応する．そこで本実験では，すべての楽曲について同一調への移調を行うことで各和音の機能が一意に定まるような枠組みを設定する．ただし，楽曲中の転調については考慮しない．つまり，ある楽曲は最初から最後まで単一の調で構成されている設定で移調を行う．本実験で使用するデータセットには，転調が起こる位置と転調後の調について記述されていない楽曲データが多数含まれているため，このような設定で楽曲の扱いを統一することとした．なお，移調の際，一般的なピアノの音域である A0 から C8 までの範囲から外れる音が生じる場合，その音はオクターブシフトすることで範囲内に収めることとした．

3.4 結果と考察

3.2 で述べた手法を用いて獲得した楽音の分散表現について，主成分分析 (Principal component analysis, PCA) を用いて 2 次元に次元圧縮することで可視化を行った (図 3.3)．この図から 2 つの特徴が読み取れる．

第一に，オクターブ違いの音がクラスタを形成している点である．オクターブ違いの音は本質的には同種の音として捉えられる．オクターブについての同値関係を特に与えていないにも関わらず，それらの音が近接する位置に配置されていることから，モデルはオクターブについての類似関係を的確に捉えることができていると考えられる．

第二に，調の構成音と非構成音，すなわちハ長調・イ短調における白鍵音と黒鍵音が図 3.3 において左右に分離して現れている点である．構成音と非構成音の間には音楽的な意味，つまりは用法に大きな違いがあると考えられる．モデルに区別を与えていないのにも関わらず，それらが特徴空間上で分離することは，音楽的意味を捉える力がモデルに備わっていることを示す結果であると考えられる．

さらに，PCA とは異なる次元圧縮のアルゴリズムである t-Distributed Stochastic Neighbor Embedding (t-SNE)[40] を用いて楽音の分散表現を二次元に圧縮した (図 3.4)．t-SNE では，空間上のデータ点間の関係を確率分布により表現し，高次元空間と低次元空間の確率分布間の差異を KL ダイバージェンスで定める．これを最小化するよう低次元空間の確率分布を学習し，そこから低次元空間の座標点群を得る．それにより，高次元空間におけるデータ点間の近傍・隣接関係を保つ次元圧縮が可能となる．PCA が線形変換であるのに対して t-SNE は非線形変換であることから，高次元空間において非線形構造をもつデータに関してより適切な低次元表現が抽出できると考えられる．学習した分散表現を t-SNE により可視化した図では， 図

3.3 で確認された二つの特徴に加え，調の構成音が C-E-G-B-D-F-A-C のように全音階的な 3 度間隔で環状に並ぶことが確認された．長調の I の和音なら C-E-G，VI の和音なら A-C-E といったように，調の任意の固有三和音は，この循環列の連続する 3 音から構成されている．このように 3 度音程は和音構成において基礎となる音程関係であり，特徴空間上の隣接関係として 3 度音程が現れることは，興味深い結果である．

以上の点から，本実験で獲得した楽音の分散表現は，オクターブについての同値関係，調の構成音と非構成音の差異，3 度音程など，音楽において重要な役割をもつ音同士の関係性を捉えていると推察できる．

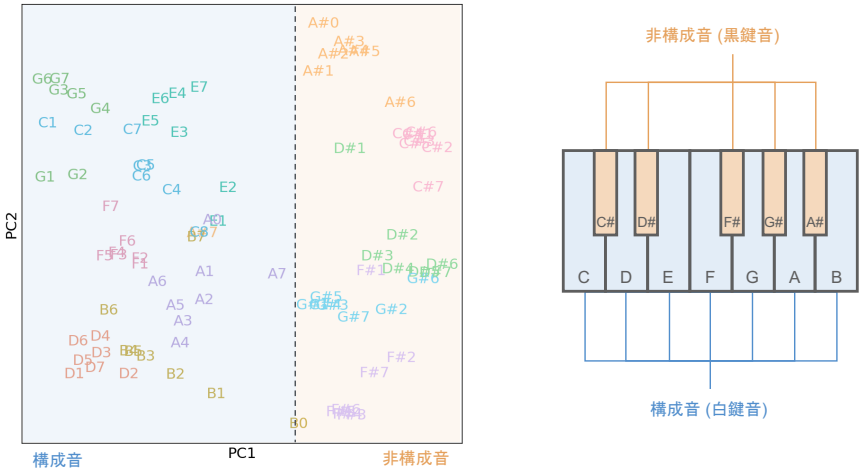
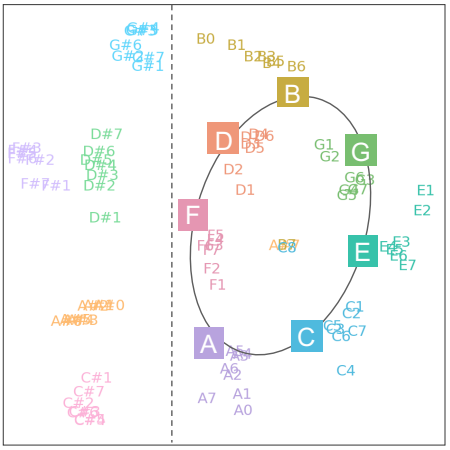
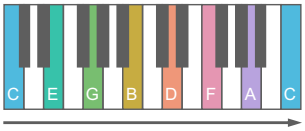


図 3.3　PCA による楽音分散表現の次元圧縮





和音	構成音
I	C-E-G
II	D-F-A
III	E-G-B
IV	F-A-C
V	G-B-D
VI	A-C-E
VII	B-D-F

図 3.4　t-SNE による楽音分散表現の次元圧縮

第 4 章

分散表現の応用

4.1 作曲家識別

4.1.1 概要

本節では，第 3 章の分散表現を用いてクラシック楽曲の作曲家識別を行う実験について述べる．クラシック音楽では，バロック・古典派・ロマン派といった時代区分があるように，楽曲が書かれた時代によって書法に違いがみられる．従来の音楽理論がそれらの違いについて定性的な説明を行う一方で，音楽情報処理では，時代や作曲家ごとの音楽的特徴を定量的に取り扱う研究が行われている．本実験では，第 3 章の分散表現を入力として用いるニューラルネットワークモデルにより，楽曲の時代区分の識別を行う．識別が正しく行われれば，時代ごとの音楽的特徴の差異を特徴空間上で表現できていると考えられ，分散表現の有用性を示すことができると考える．

4.1.2 手法

本実験では，Attention 機構 [41] を導入した双方向 LSTM[42] を用いて，和音系列から楽曲の時代区分を識別するモデルを構築する．

順序が意味をもつ時系列データを扱う基本的なニューラルネットワークモデルとして Recurrent Neural Network (RNN)[43] が知られている．RNN では，時系列データにおける現時点までの要素から未来の要素を推論するモデルの学習を行う．RNN では，入力の系列が長くなると学習時に勾配消失を起こしやすくなり，系列内の遠く離れた要素の間に存在する依存関係の学習が困難となる場合がある [44]．これは誤差逆伝播法により損失の勾配を計算する際，系列の長さの分だけ重みを乗じる回数が増加し，勾配の大きさが指数関数的に減少することが原因である．RNN のネットワークに新たなユニットを追加することで勾配消失の問題に対処したモデルが Long Short-Term Memory (LSTM)[45] である．さらに，入力系列を未来から過去に向かって逆方向に学習するユニットを LSTM に導入し，順方向のユニットと連結するモデルを双方向 LSTM[42] とよぶ．

Attention 機構 [41] とは，入力系列の各要素に対して重み付けをする手法である．このとき，入力系列のうちタスクを解くうえで重要な情報をもつ要素に対して大きな重みを与えられ

るよう，ネットワークの学習が行われる．

本実験では，双方向 LSTM と Attention 機構を組み合わせ，和音系列から楽曲の時代区分の識別を行うモデルを構築する．モデル構成の概要を図 4.1 に示す．モデルへの入力として楽曲中に出現する連続した和音の系列を用いる．以後，モデルの入力に用いる和音系列を入力ユニットとよぶ．本実験では，入力ユニットの系列長を 64 に設定する．モデル内部では，まず，入力ユニットに含まれる各和音の構成音に第 3 章で学習した分散表現を割り当てることで，各和音を楽音ベクトルの集合に変換する．次に，Attention 機構を用いて各和音の構成音ベクトルに重み付けし，それらを足し合わせることで各和音のベクトルとする．これにより入力ユニットをベクトル系列として扱うことが可能となる．得られたベクトル系列を双方向 LSTM の入力とし，Attention 機構を用いて LSTM 層の各時点における隠れ状態に重み付けを行う．重み付けした隠れ状態を足し合わせることで得られる特徴ベクトルに対して線形変換を行い，softmax 関数による正規化を行う．これにより，入力ユニットがバロック・古典派・ロマン派の各クラスに属する確率を出力する．

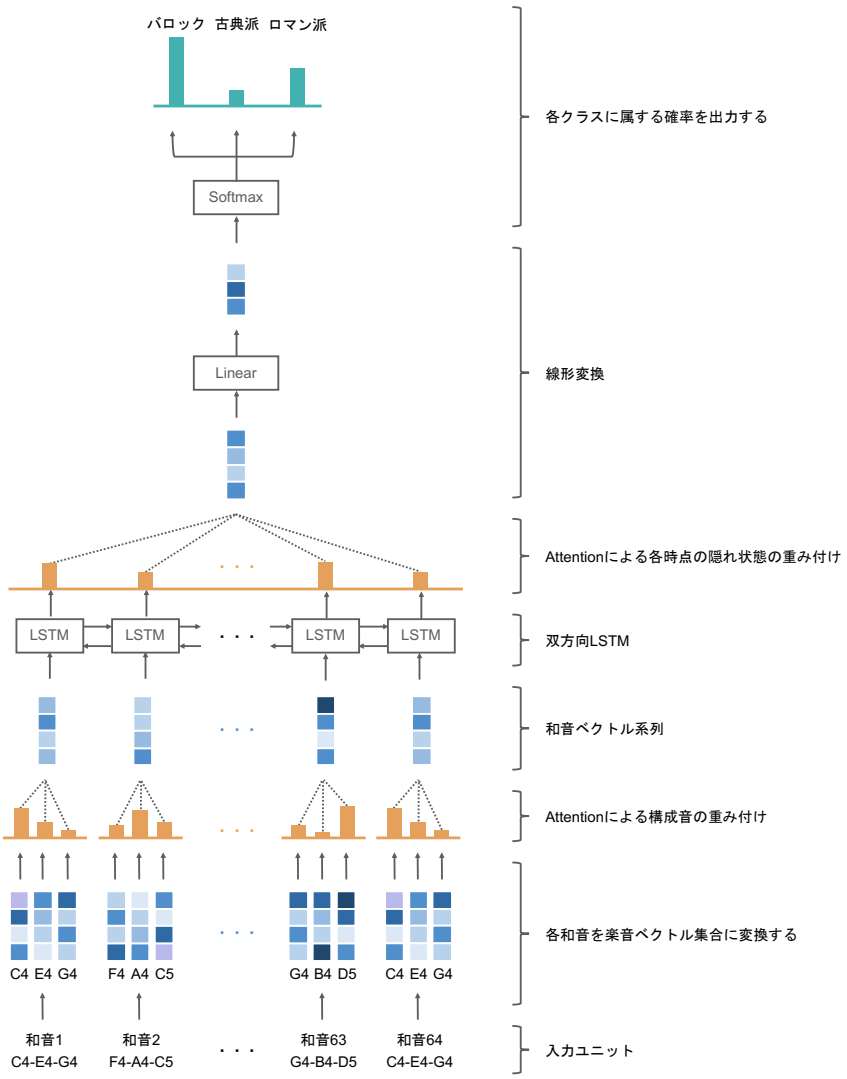


図 4.1　和音系列の時代区分識別モデル

楽曲データから入力ユニットを作成する方法を以下に示す．第 3 章の実験と同様の方法で楽曲データを和音系列に変換し，それを元に入力ユニットを作成する．まず，楽曲の 1 番目の和音から 64 番目の和音までの和音系列を切り出す．次に，和音 32 個分後ろにずらし，33 番目から 96 番目までの和音系列を切り出す．さらに 32 個分後ろにずらし，65 番目から 128 番目までの和音系列を切り出す．このように，楽曲から窓幅 64，ステップ幅 32 で和音系列を切り出す操作を楽曲中の最後の和音に達するまで行う．この方法で楽曲から切り出した各和音系列を入力ユニットとして用いる．なお，楽曲から和音系列を切り出す際，楽曲の最後に系列長 64 未満の和音系列が生じる場合，その和音系列は入力ユニット集合から除外する．

モデルの評価は 5 分割交差検証により行う (図 4.2)．まず，実験に用いる楽曲集合において，バロック・古典派・ロマン派の各クラスをそれぞれ 5 等分する．各クラスにおいて，5 分割したうちの 1 つのデータ群をテストデータとし，残る 4 つのデータ群を学習データとする．各クラスの学習データとテストデータをそれぞれ統合することで，楽曲集合全体を学習データとテストデータに分割する．各クラスにおいてテストデータとして用いるデータ群を変えながら計 5 パターンの学習データ・テストデータを作成し，それぞれモデルの学習・評価を行う．5 回分の評価結果を平均することでモデルの評価を行う．

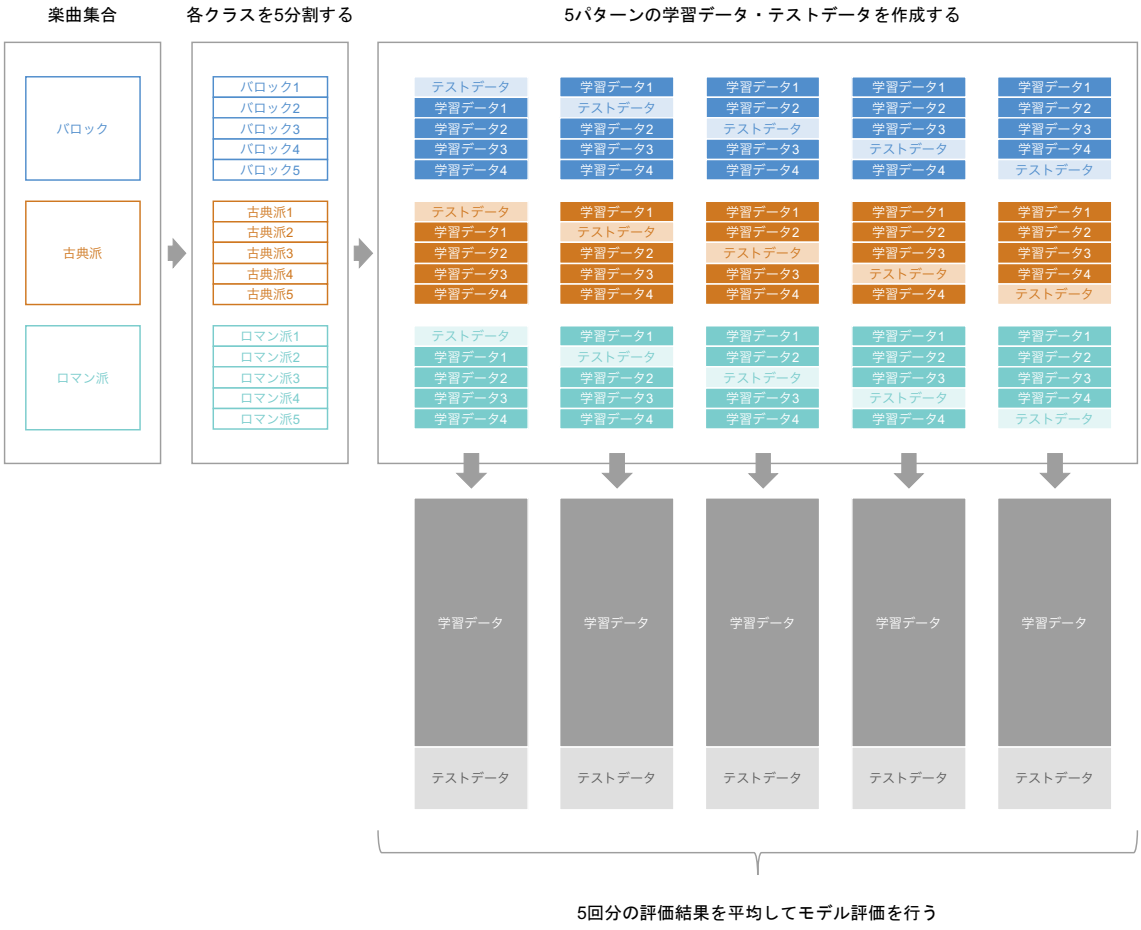


図 4.2　5 分割交差検証によるモデル評価

4.1.3 使用データ

第 3 章で使用した Kunst der Fuge[39] よりダウンロードしたデータを使用する．本実験では，Kunst der Fuge に含まれるクラシック音楽の作曲家 13 人の楽曲 2,396 曲分（表 4.1）の MIDI ファイルを用いる．なお，複数の楽章に分かれた楽曲に関しては，各楽章を一つの楽曲として数える．また，第 3 章と同様に，すべての楽曲について長調の曲はハ長調，短調の曲はイ短調に移調して実験に用いる．楽曲中の転調を考慮しない点についても第 3 章と同様である．

表 4.1　使用データ

時代区分	作曲家	曲数	入力ユニット数
バロック 4 作曲家・1,028 曲	A. L. Vivaldi	20	23,997
	G. P. Telemann	49	
	G. F. Händel	238	
	J. S. Bach	721	
古典派 2 作曲家・683 曲	F. J. Haydn	462	25,063
	W. A. Mozart	221	
ロマン派 7 作曲家・685 曲	M. Mendelssohn	46	24,988
	F. F. Chopin	107	
	R. Schumann	67	
	J. Brahms	127	
	A. L. Dvořák	94	
	P. I. Tchaikovsky	175	
	G. U. Fauré	69	
	合計	2,396	74,048

4.1.4 結果と考察

時代識別において，各クラスの入力ユニットが実際にどのクラスと識別されたかを表す混同行列を図 4.3 に示す．図中の数値は，5 分割交差検証の各回において，識別された入力ユニット数の平均値である．また，この識別結果から算出される性能評価値を表 4.2 に示す．モデルによる識別精度は約 76% であった．クラスごとの結果に着目すると，時代が下るほど再現率が低下し，適合率は上昇していることがわかる．混同行列の結果をみると，ロマン派の楽曲を古典派，古典派の楽曲をバロックであると誤識別する割合が特に大きく，次いでロマン派の楽曲をバロックであると誤識別する割合が大きい．このことから，後続する時代の楽曲を先行する時代の楽曲であると誤識別するケースが多いことがわかる．これらの結果は，音楽史的観点から妥当であると考えられる．クラシック音楽において，後続時代の楽曲は先行時代の作曲様

式を踏襲・応用して作られてきた．したがって，先行時代に編み出された作曲技法が後続時代の楽曲で用いられることは一般的である．このため，後続時代の楽曲には先行時代の楽曲と明確に区別することが難しいものがあり，そのような楽曲を誤識別するケースが多いと考えられる．

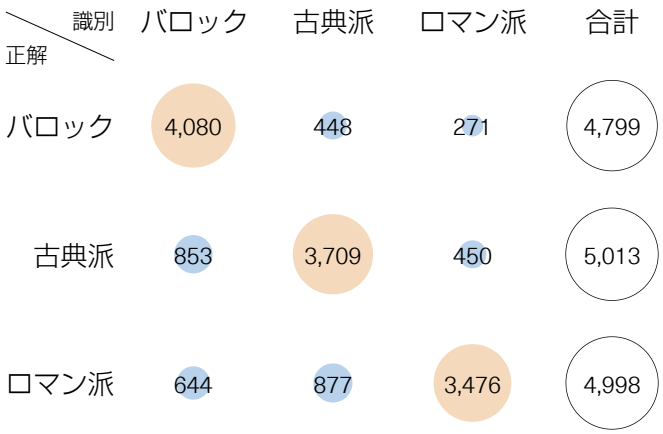


図 4.3　時代識別の混同行列

表 4.2　時代識別モデルの性能評価

クラス	精度	適合率	再現率	F 値
バロック	—	73.0%	85.0%	78.4%
古典派	—	74.1%	74.0%	73.7%
ロマン派	—	82.7%	69.6%	75.5%
平均	76.0%	76.6%	76.2%	75.9%

4.2 楽曲間類似度算出

4.2.1 概要

本節では，第 3 章で得られた分散表現を活用した楽曲間類似度計算の実験について述べる．自然言語処理における分散表現学習手法の発展に伴い，単語の分散表現に基づいて文と文の類似度を計算する手法が提案されている．Kusner らによって 2015 年に提案された Word Mover's Distance (WMD)[30] はその一つである．WMD では，線形計画法の輸送問題に基づいて 2 つの分布間の距離を測定する Earth Mover's Distance (EMD)[31] を用いる．本実験では，第 3 章で学習した分散表現を活用し，WMD を楽曲に適用することで楽曲間の類似度を計算する方法を提案する．本実験の手法を用いて得られる楽曲間類似度をもとに楽曲のクラスタ分析を行い，得られた結果について音楽学の知見に基づく考察を行うことで手法の妥当性について議論する．

4.2.2 手法

WMD では文 x から文 y への単語の輸送問題を考えることで文間の距離を計算する．これを楽曲 x から楽曲 y への和音の輸送問題に置き換えて考えることで，楽曲間距離の計算に WMD を適用することができる．

WMD を適用するため，第 3 章の実験と同様の方法で楽曲データを和音系列に変換する．各和音の構成音に第 3 章で学習した分散表現を割り当て，それら構成音ベクトルの相加平均を各和音のベクトル表現とする．和音の輸送問題において，和音 i から和音 j への輸送コストを 2 つの和音ベクトル間のユークリッド距離と定義し，$c(i, j)$ で表す．楽曲は正規化された Bag-of-Words (nBOW) ベクトル $\boldsymbol{d}$ で表現する．ここで，nBOW ベクトルの次元数は楽曲中に出現する和音の種類の数となり，ベクトルの各要素は対応する和音の出現頻度となる．2 つの楽曲の nBOW 表現を $\boldsymbol{d}, \boldsymbol{d}'$ とし，$\boldsymbol{d}$ 内の各和音 i から $\boldsymbol{d}'$ 内の各和音 j への輸送量を配列した行列を $\boldsymbol{T}_{ij}$ とする．和音の種類の数を n とすると，以下の式で定義される WMD の最適化問題として楽曲間距離を計算することができる．

$$\min_{\boldsymbol{T}_{ij} \geq 0} \sum_{i,j=1}^{n} \boldsymbol{T}_{ij} c\,(i, j) \tag{4.1}$$

$$\text{subject to:} \quad \sum_{j=1}^{n} \boldsymbol{T}_{ij} = \boldsymbol{d}_i \qquad (1 \leq i \leq n) \tag{4.2}$$

$$\sum_{i=1}^{n} \boldsymbol{T}_{ij} = \boldsymbol{d}'_j \qquad (1 \leq i \leq n) \tag{4.3}$$

この最適化問題を解くことによって，和音同士の音楽的関係性を反映した楽曲間の距離を計算することが可能になると考えられる．

さらに，WMD により計算される楽曲間距離をもとに楽曲のクラスタ分析を行う．まず，使用する楽曲の全ての組み合わせについて WMD により距離を算出し，それらを配列することで距離行列を作成する．得られた距離行列を用いて凝集型階層的クラスタリングの手法である

ウォード法 [46] を用いてクラスタ分析を行う．

凝集型階層的クラスタリングでは，全てのクラスタ対の距離を計算し，最も近いクラスタを併合する．併合後の新たなクラスタを含む全てのクラスタ対について距離を再計算し，クラスタの併合を行う．全体が一つのクラスになるまでこれを繰り返し行う．凝集型階層的クラスタリングには，クラスタ間の距離を求める方法の違いによって様々な手法が存在するが，本実験ではウォード法によるクラスタ分析を行う．ウォード法では，併合後のクラスタの分散と併合前の各クラスタの分散の和との差をクラスタ間の距離と定義し，これを最小化するようクラスタの併合を行う．

4.2.3 使用データ

クラシック音楽の MIDI ファイルを提供する Web サイトである Kunst der Fuge[39], KernScores[47] よりダウンロードした楽曲データを用いる．クラシック音楽の作曲家 4 人のピアノ曲 147 曲分 (表 4.3) の MIDI ファイルを用いる．バロックの楽曲として J. S. Bach の平均律クラヴィーア曲集，古典派の楽曲として W. A. Mozart のピアノソナタ，ロマン派以降の楽曲として F. F. Chopin の練習曲，マズルカ，A. Scriabin の練習曲，マズルカを使用する．Scriabin は時期によって作風が異なるが，本実験では Chopin の影響を強く受けていたとされる初期の作品を用いる．なお，複数の楽章に分かれた楽曲に関しては，各楽章を一つの楽曲として数える．本研究で行なったこれまでの実験と同様に，すべての楽曲について長調の曲はハ長調，短調の曲はイ短調に移調して用いる．楽曲中の転調を考慮しない点についても同様である．

表 4.3　使用楽曲データ

作曲家	楽曲		使用曲数 (楽章数)
J. S. Bach	The Well-Tempered Clavier	Book I	23
		Book II	22
W. A. Mozart	Piano Sonata	K. 279	3
		K. 280	3
		K. 281	1
		K. 282	3
		K. 283	3
		K. 284	1
		K. 309	2
		K. 310	2
		K. 311	2
		K. 330	3
		K. 332	2
		K. 333	1
		K. 457	3
		K. 545	3
		K. 570	3
		K. 576	3
F. F. Chopin	Etude	Op. 10	8
		Op. 25	2
	Mazurka	Op. 6	3
		Op. 7	3
		Op. 17	4
		Op. 24	3
		Op. 30	2
		Op. 33	2
		Op. 41	3
		Op. 50	3
		Op. 56	3
		Op. 59	3
		Op. 63	2
		Op. 67	1
A. Scriabin	Etude	Op. 8	10
		Op. 42	3
	Mazurka	Op. 3	9

4.2.4 結果と考察

ウォード法によるクラスタ分析によって得られた樹形図を図 4.6 に示す．クラスタ分析の結果，データは大きく 3 つのクラスタに分けられることがわかった．3 つのクラスタそれぞれに含まれる各楽曲の作曲家 (Bach, Mozart, Chopin, Scriabin) をみると，3 つのクラスタはそれぞれ Bach のクラスタ，Mozart のクラスタ，Chopin と Scriabin のクラスタに概ね分かれていることが確認できた．これら 3 つのクラスタはそれぞれバロック，古典派，ロマン派以降という時代区分に一致する．このことから，WMD を用いて得られる楽曲間距離には，楽曲の時代区分による音楽的特徴の差異が反映されていると考えられる．また，本実験で用いた Scriabin の初期の作品は Chopin の影響を強く受けているとされる [48]．このことから，時代区分の違いとは別に，Scriabin の初期作品にみられる Chopin の影響のような非自明な類似構造を捉えている可能性も考えられる．以後，Bach のクラスタを Ba，Mozart のクラスタを Mo，ロマン派以降のクラスタを Ro と表現する．

Ro クラスタに内包される楽曲は，Ro クラスタ内部でさらに 3 つのクラスタに分けられる．3 つのクラスタそれぞれに含まれる各楽曲について，作曲家と調性 (長調か短調か) の組み合わせによって分類した．その結果，3 つのクラスタはそれぞれ，Scriabin の楽曲を多く含む短調クラスタ (Ro-m1)，Chopin の楽曲のみで構成される小規模な短調クラスタ (Ro-m2)，長調クラスタ (Ro-M) として概ね解釈できる．また，Ba クラスタに内包される楽曲は Ba クラスタ内部でさらに 2 つのクラスタに分けられる．各クラスタに含まれる各楽曲について作曲家と調性の組み合わせによる分類を行った結果，長調クラスタ (Ba-M) と短調クラスタ (Ba-m) にはっきりと分離されることが確認できた．Mo クラスタについては他のクラスタと異なり，長調と短調で大きく分かれているわけではない．しかし，Mo クラスタ内では Mozart の短調の楽曲が凝集し小規模なクラスタを形成していることが確認できる．この点から，Mozart の短調作品は少数であるものの，長調作品との差異については捉えることができていると考えられる．

これらのことから，WMD を用いて得られる楽曲間距離について，大局的には時代区分の違いが影響し，各時代区分においては調性の違いが大きく影響すると考えられる．

本項で述べた時代区分や調性により説明されるクラスタのうち，ロマン派以降の短調クラスタ (Ro-m1)，ロマン派以降の長調クラスタ (Ro-M)，Bach の長調クラスタ (Ba-M) の 3 つのクラスタについて，それらに内包されるより小さなクラスタを分析し考察を行う．

Ro-m1 クラスタでは，Scriabin の楽曲のみが凝集した 2 つの小規模なクラスタが確認できる．このことから，作曲家の特徴を捉えることができていると考えられる．

Ro-M クラスタに内包される楽曲を 2 つのクラスタに分けると，一方は Scriabin の楽曲や短調の楽曲を複数含むクラスタとなり，もう一方は Mozart と Chopin の長調の楽曲で構成される小規模なクラスタとなった．前者のクラスタは作曲家や調性の区別が曖昧になっていることから，後者のクラスタと比較して多様かつ複雑な和声を使用した楽曲群であることが推察される．

Ba-M クラスタには，Scriabin の練習曲 Op. 8 の No. 6 が含まれている．クラスタを構成する楽曲の大半を占める Bach はバロック時代の作曲家であるのに対して，Scriabin は後期ロマ

ン派から近代にかけての作曲家であり，時代区分の観点ではかけ離れている．また，クラスタ内にはこの楽曲の他にロマン派以降の楽曲は含まれておらず，異質である．この理由について，楽曲中で使用される音程関係の観点から考察した．Op. 8 の No. 6 の特徴的な点として，6 度の重音が曲中で多く出現する点があげられる (図 4.4)．一方，本実験で使用した Bach の平均律クラヴィーア曲集に含まれる各楽曲は前奏曲とフーガで構成されており，対位法的手法が駆使される．対位法とは，独立に進行する複数の旋律を調和させながら重ねる技法であり，1 度，8 度，完全 5 度，3 度，6 度の協和音を基礎として旋律が重ね合わせられる．協和音の中では，1 度と 8 度の音程は各声部の独立性を損なうため避けられる場合があり，5 度の音程に関しても慣習により使用を制限する規則が存在する [49]．その結果として対位法的手法を用いる楽曲では 3 度や 6 度の音程による進行が出現しやすくなる．図 4.5 に示した譜例は Bach の平均律クラヴィーア曲集第 2 巻の第 5 番から抜粋した 6 度音程による進行の例である．この曲は Scriabin の練習曲 Op. 8 の No. 6 と最も近いクラスタを構成する楽曲の一つである．これらのことから，対位法において頻出する 6 度音程による進行と Op. 8 の No. 6 で多用される 6 度の重音進行との類似性をモデルが捉えている可能性が考えられる．

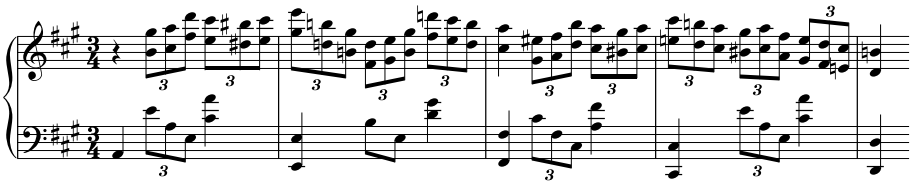


図 4.4　Scriabin: 12 Etudes Op. 8, No. 6 より

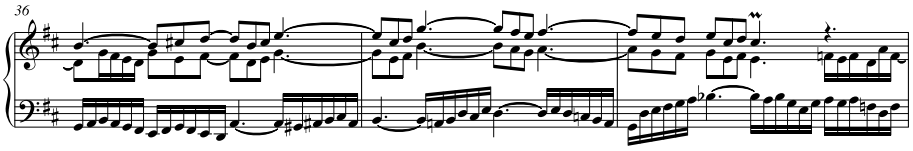


図 4.5　Bach: The Well-Tempered Clavier Book II, No. 5 より

以上のことから，WMD を用いて得られる楽曲間距離は，楽曲の時代区分や調性の違いのみならず，作曲家ごとの特徴や，それらで説明できない個々の楽曲の特徴までも反映していることが示唆される．

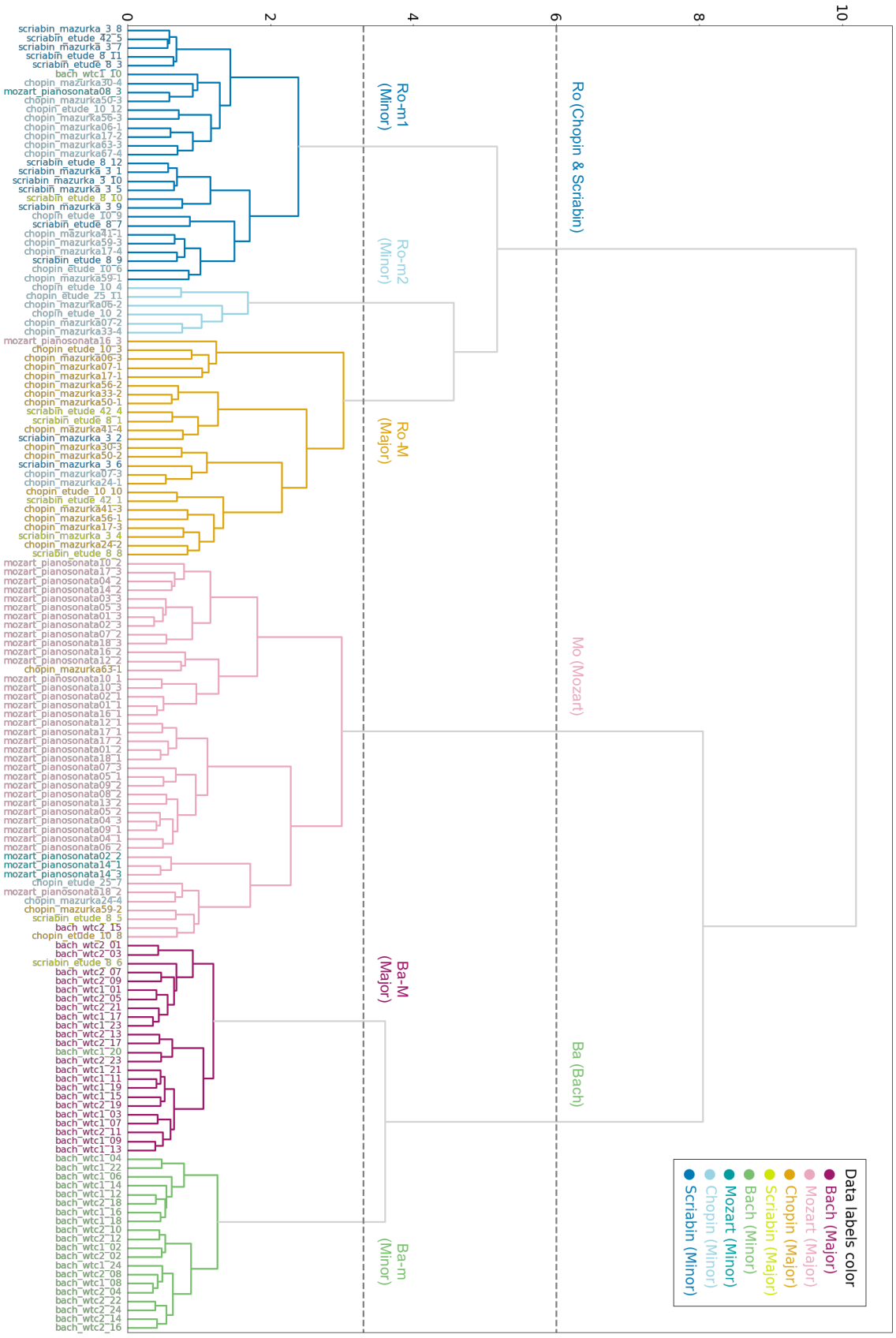


図 4.6　ウォード法による階層クラスタ分析の結果

第 5 章

まとめと今後の課題

5.1 まとめ

音楽には形式的に記述されない暗黙知が多く含まれており，その構造や意味をモデル化することが難しい．これは音楽情報処理の諸課題に共通して関わる問題である．一方，自然言語にも類似の性質があり，自然言語処理ではこの問題に対処するため確率論や統計学に基づく方法が用いられている．特に近年では，単語の分散表現を用いた手法が大きな成果をあげている．自然言語と音楽の類似性から，音楽の構成要素である和音に分散表現を適用することで音楽をモデル化する試みもある．一方，音楽理論においては音階上の各楽音の性格の違いが述べられていることから，楽音を単位とした分散表現により音楽をモデル化する手法も考えられる．しかし，これまでに楽音を単位とした分散表現学習を行う研究はほとんど行われていない．

本研究では，音楽の最も基本的な構成要素である楽音に対して分散表現を適用することで，曖昧性をもつ音楽の構造や意味をモデル化することを目指す．自然言語処理における代表的な分散表現学習手法である word2vec を応用し，各楽音の音楽的性格を捉えたベクトル空間の獲得を行った．さらに楽音の分散表現を元により大きな単位である和音や楽曲の分散表現を合成し，それらを用いてクラシック音楽を対象とした楽曲の時代識別と楽曲間類似度の計算を行った．時代識別では，楽音の分散表現を入力とするニューラルネットワークにより，楽曲をバロック，古典派，ロマン派の 3 クラスに分類する．楽曲間類似度では，単語の分散表現を用いた文間類似度の計算手法である Word Mover's Distance (WMD) を応用し，楽曲間の類似度を計算する．これらの実験結果について音楽学的な観点から妥当性を考察し，分散表現を活用する手法の有用性や応用可能性について議論する．

学習した楽音の分散表現を 2 次元圧縮し可視化した結果，(i) オクターブ違いの音がクラスタを形成し，(ii) 調の構成音クラスタと非構成音クラスタが左右に分離し，(iii) 調の構成音クラスタが全音階的な 3 度間隔で環状に並ぶことが確認された．時代識別では，クラシック音楽の楽曲 2,396 曲を対象とした 3 クラス分類を行い，約 76% の識別精度を示した．WMD を用いて得られた楽曲間距離を配列した距離行列に対してウォード法による階層クラスタ分析を行なった結果，(i) 楽曲の時代区分の違いによって大きく 3 つのクラスタが形成され，(ii) 各時代クラスタ内部では長調/短調の違いにより中程度のクラスタが形成されることを確認し，(iii) さらに下の階層では作曲家や個々の楽曲の特徴を捉えていることを示唆する結果が得ら

れた．

これらの結果には音楽理論の観点から妥当性があると考えられる．このことから形式的に記述することが難しい音楽の構造や意味をモデル化するための方法として，楽音の分散表現を活用するアプローチは有効であると考えられる．楽曲間類似度については楽曲の時代区分や調性の違いのみならず，作曲家ごとの特徴やそれらで説明できない個々の楽曲の特徴までも捉えていることが示唆されており，音楽情報検索などの課題に対しても応用可能であると考えられる．さらに，これまで音楽学において対比されることのなかった楽曲間の類似性についてモデルから知見を得られる可能性もあり，音楽学においても有用なツールとして利用できる可能性がある．

5.2 今後の課題

本研究では，すべての楽曲について同一調への移調を行うことで各和音の機能が一意に定まるような枠組みを設定した．ただし，本研究で使用したデータセットには，転調が起こる位置と転調後の調について記述されていない楽曲データが多数含まれているため，楽曲中の転調については考慮せず移調を行うこととした．したがって，部分的に実態と異なるデータを使用していることとなり，その点が結果に影響を与えている可能性がある．本研究の手法についてより正確な分析を行うためには，転調が起こる位置と調について正確にアノテーションされたデータセットを作成し用いる必要がある．

音楽理論においては，同時に鳴っていても和音の構成音とされない非和声音や，和音の構成音を時間的にずらして配置する分散和音などがある．しかし，本研究では楽曲中で同時に鳴る楽音の集合を和音と定義したため，非和声音や分散和音などについては考慮していない．より音楽理論に則した枠組を目指すならば，非和声音や分散和音などについても考慮する必要がある．

本研究では，分散表現学習の手法として word2vec を用いた．一方，自然言語処理においては word2vec の他にも様々な手法が提案されている．しかし，本研究ではそれらの手法については十分に検討できていない．

本研究では，学習した楽音の分散表現を合成することで和音の分散表現を獲得した．一方，先行研究では和音系列から和音の分散表現を学習する事例が多くある．現状では，先行研究の手法と本研究の手法について，得られる分散表現の性能を十分に比較することができていない．和音の意味は明確に定義されるものではないため，自然言語処理における単語の分散表現と同様の評価方法を使用するのは困難である．これに関しては，分散表現の性能を評価する方法も含めて今後検討していく必要がある．

本研究では，WMD による楽曲間の距離行列を用いたクラスタリングについて，楽曲の時代区分や調性，作曲家などのラベルとクラスタを対応づけて解釈することで妥当性の考察を行った．しかし，実際には時代区分や調性などのラベルの違いでは説明することができない非自明な類似構造を的確に捉えられるかどうかが重要である．Scriabin の初期作品にみられる Chopin の影響や 6 度重音と対位法に関する考察はその例といえるが，これらだけでは分析の質・量ともに不十分である．提案手法の妥当性をより適切に評価するための方法として，音楽

学的分析と照らし合わせる方法が考えられる．幅広い時代・地域の作曲家の楽曲を対象に音楽学に基づく比較・分析を行い，楽曲間の類似関係に関する定性的な説明を得た上で，計算により得られた結果と照らし合わせ考察する必要がある．一方で，これまで音楽学において対比されることのなかった楽曲間の類似性についてモデルから知見を得られる可能性もある．その点でも，本研究で提案する楽曲間類似度の尺度と音楽学に基づく分析手法を相互に組み合わせる分析は有用であると考えられる．

謝辞

本論文を執筆するにあたり，研究指導教員である平賀譲先生，副研究指導教員である寺澤洋子先生には，温かなご指導ご鞭撻を賜りました．心より感謝申し上げます．

人と音の情報学研究室の仲間である小池栄美さん，社本和磨さん，中川稜介さん，若狭健太さん，池田周平くん，河合優理子さん，川島涼太くん，宮澤響くん，大中悠生くん，小島直くん，相馬翔太くん，初見佳那子さん，松本悠路くん，山本雄也くん，都築陵佑くん，中岡想太郎くん，三井颯人くん，水野真由美さん，森山大地くんとは，互いに助け合いながら研究を進めることができました．感謝いたします．最後に，大学生活，大学院生活を遠方から支えてくれた家族に深く感謝いたします．

参考文献

[1] 東条敏，平田圭二：音楽・数学・言語――情報科学が拓く音楽の地平，近代科学社 (2017).

[2] 後藤真孝：特集――新しい○○情報学――音楽情報学，情報処理，Vol. 51，No. 6，pp. 661–668 (2010).

[3] 平田圭二，東条敏，浜中雅俊，平賀譲：道しるべ――計算の視点から音楽の構造を眺めてみると――第 1 回――計算論的音楽理論について，情報処理，Vol. 49，No. 7，pp. 824–830 (2008).

[4] 吉井和佳：特集 B――音楽情報処理・音楽信号処理の最前線――2 章――音楽と統計的記号処理，情報処理メディア学会誌，Vol. 71，No. 4，pp. 457–461 (2017).

[5] Haruto Takeda, Naoki Saito, Tomoshi Otsuki, Mitsuru Nakai, Hiroshi Shimodaira, and Shigeki Sagayama, "Hidden Markov Model for Automatic Transcription of MIDI Signals," In *Proceedings of 2002 IEEE Workshop on Multimedia Signal Processing*, pp. 428–431 (2002).

[6] Kyogu Lee and Malcolm Slaney, "Automatic Chord Recognition from Audio Using an HMM with Supervised Learning," In *Proceedings of the 7th International Conference on Music Information Retrieval*, pp. 133–137 (2006).

[7] Kazuyoshi Yoshii and Masataka Goto, "A Vocabulary-Free Infinity-Gram Model for Nonparametric Bayesian Chord Progression Analysis," In *Proceedings of the 12th International Conference on Music Information Retrieval*, pp. 645–650 (2011).

[8] Masato Tsuchiya, Kazuki Ochiai, Hirokazu Kameoka, and Shigeki Sagayama, "Probabilistic Model of Two-Dimensional Rhythm Tree Structure Representation for Automatic Transcription of Polyphonic MIDI Signals," *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1315–1320 (2013).

[9] G. E. Hinton, James L. McClelland, and D. E. Rumelhart, "Distributed Representations," *Parallel Distributed Processing: Explorations in the Microstructure of Cognition—Volume 1: Foundations*, pp. 77–109 (1986).

[10] 岡崎直観：言語処理における分散表現学習のフロンティア――特集――ニューラルネットワーク研究のフロンティア――言語処理における分散表現学習のフロンティア，人工知能，Vol. 31，No. 2，pp. 189–201 (2016).

[11] Sephora Madjiheurem, Lizhen Qu, and Christian Walder, "Chord2Vec: Learning Musical Chord Embeddings," In *Proceedings of the Constructive Machine Learning 2016*, pp. 1–5 (2016).

[12] Cheng-Zhi Anna Huang, David Duvenaud, and Krzyszof Z. Gajos, “ChordRipple: Recommending Chords to Help Novice Composers Go Beyond the Ordinary,” In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pp. 241–250, ACM (2016).

[13] Dorien Herremans and Ching-Hua Chuan, “Modeling Musical Context Using Word2vec,” In *Proceedings of the First International Workshop on Deep Learning and Music*, pp. 11–18 (2017).

[14] Ali Nikrang, David R. W. Sears, and Gerhard Widmer, “Automatic Estimation of Harmonic Tension by Distributed Representation of Chords,” *Music Technology with Swing*, pp. 23–34, Springer International Publishing (2018).

[15] 塚本康太，饗庭絵里子，南泰浩：単語埋め込みを利用した和音進行分析，情報処理学会研究報告，Vol. 2019-MUS-122，No. 9 (2019).

[16] 石田颯人，木村昌臣：度数表記と Chord2Vec を利用した楽曲類似度指標の提案，情報処理学会研究報告，Vol. 2019-MUS-123，No. 21 (2019).

[17] George A. Miller, “WordNet: A Lexical Database for English,” *Communications of the ACM*, Vol. 38, No. 11, pp. 39–41 (1995).

[18] Zellig S. Harris, “Distributional Structure,” *WORD*, Vol. 10, No. 2-3, pp. 146–462 (1954).

[19] Marco Baroni, Georgiana Dinu, and Germán Kruszewski, “Don’t count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors,” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Vol. 1, pp. 238–247 (2014).

[20] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman, “Indexing by Latent Semantic Analysis,”textitJournal of the American Society for Information Science, Vol. 41, No. 6, pp. 391–407 (1990).

[21] Thomas Hofmann, “Probabilistic Latent Semantic Indexing,” In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 50–57 (1999).

[22] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022 (2003).

[23] Tomáš Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean, “Distributed Representations of Words and Phrases and their Compositionality,” In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, Vol. 2, pp. 3111–3119, Curran Associates, Inc. (2013).

[24] Tomáš Mikolov, Wen-tau Yih, and Geoffrey Zweig, “Linguistic Regularities in Continuous Space Word Representations,” In *Proceedings of the 2013 North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 746–751, Association for Computational Linguistics (2013).

[25] Jeffrey Pennington, Richard Socher, and Christopher D. Manning, “GloVe: Global

Vectors for Word Representation," In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1532–1543, Association for Computational Linguistics (2014).

[26] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomáš Mikolov, "Enriching Word Vectors with Subword Information," *Transactions of the Association for Computational Linguistics*, Vol. 5, pp. 135–146 (2017).

[27] Alexander Budanitsky and Graeme Hirst, "Evaluating WordNet-based Measures of Lexical Semantic Relatedness," *Computational Linguistics*, Vol. 32, No. 1, pp. 13–47 (2006).

[28] Lev Finkelstein, Evgeniy Gabrilovich, Yossi Matias, Ehud Rivlin, Zach Solan, Gadi Wolfman, and Eytan Ruppin, "Placing Search in Context: The Concept Revisited," *ACM Transactions on Information Systems*, Vol. 20, No. 1, pp. 116–131 (2002).

[29] Felix Hill, Roi Reichart, and Anna Korhonen, "SimLex-999: Evaluating Semantic Models With (Genuine) Similarity Estimation," *Computational Linguistics*, Vol. 41, No. 4, pp. 665–695 (2015).

[30] Matt J. Kusner, Yu Sun, Nicholas I. Kolkin, and Kilian Q. Weinberger, "From Word Embeddings To Document Distances," In *Proceedings of the 32nd International Conference on Machine Learning*, Vol. 37, pp. 957–966 (2015).

[31] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas, "The Earth Mover's Distance as a Metric for Image Retrival," *International Journal of Computer Vision*, Vol. 40, No. 2, pp. 99–121 (2000).

[32] 石桁真礼生，丸田昭三，金光威和雄，末吉保雄，飯田隆，飯沼信義：楽典――理論と実習，音楽之友社 (1965).

[33] 青島広志：究極の楽典――最高の知識を得るために，全音楽譜出版社 (2009).

[34] 島岡譲：和声のしくみ・楽曲のしくみ――4 声体・キーボード・楽式・作曲を総合的に学ぶために，音楽之友社 (2006).

[35] 島岡譲，丸田昭三，小林秀雄，池内友次郎，長谷川良夫，石桁真礼生，松本民之助，柏木俊夫，三善晃，末吉保雄，佐藤眞：和声――理論と実習――I，音楽之友社 (1964).

[36] Emanuele Pollastri and Giuliano Simoncelli, "Classification of Melodies by Composer with Hidden Markov Models," In *Proceedings of the First International Conference on WEB Delivering of Music*, pp. 88–95, IEEE Computer Society (2001).

[37] ヤン・ラルー，大宮眞琴：スタイル・アナリシス――1――綜合的様式分析――方法と範例，音楽之友社 (1988).

[38] 長谷川隆，西本卓也，小野順貴，嵯峨山茂樹：楽譜情報からの作曲家らしさ認識のための音楽特徴量の提案，情報処理学会論文誌，Vol. 53，No. 3，pp. 1204–1215 (2012).

[39] *Kunst der Fuge* [Online]. Available: `http://www.kunstderfuge.com/`

[40] Laurens van der Maaten and Geoffrey Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, Vol. 9, pp. 2579–2605 (2008).

[41] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uskoreit, Llion Jones, Aidan N.

Gomez, and Łukasz Kaiser, “Attention Is All You Need,” In *Proceedings of the 31st Conference on Neural Information Processing System*, pp. 5998–6008, Curran Associates Inc. (2017).

[42] Mike Schuster and Kuldip K. Paliwal, “Bidirectional Recurrent Neural Networks,” *IEEE Transactions on Signal Processing*, Vol. 45, No. 11, pp. 2673–2681 (1997).

[43] Jeffrey L. Elman, “Finding Structure in Time,” *Cognitive Science*, Vol. 14, No. 2, pp. 179–211 (1990).

[44] Yoshua Bengio, Patrice Simard, and Paolo Frasconi: “Learning Long-Term Dependencies with Gradient Descent is Difficult,” *IEEE Transactions on Neural Networks*, Vol. 5, No. 2, pp. 157–166, IEEE (1994).

[45] Sepp Hochreiter and Jürgen Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780, MIT Press (1997).

[46] Joe H. Ward Jr., “Hierarchical Grouping to Optimize an Objective Function,” *Journal of the American Statistical Association*, Vol. 58, No. 301, pp. 236–244 (1963).

[47] *KernScores* [Online]. Available: `http://kern.ccarh.org/`

[48] 志賀眞知子：スクリャービン――ピアノ音楽語法の変遷――小品からのアプローチ，大阪芸術大学紀要，No. 25，pp. 54–66 (2002).

[49] アルノルト・シェーンベルク，レナード・スタイン（編），山縣茂太郎（訳），鴫原真一（訳）：シェーンベルク――対位法入門，音楽之友社 (1978).