

単語埋め込みを利用した和音進行分析

塚本康太^{†1} 饗庭絵里子^{†1,2} 南泰浩^{†1,3}

概要：本研究では、単語埋め込みを利用した和音のベクトル化により、和音進行の規則をより明確なものとする。和音の中には、トニックかドミナントなのか判別しづらいものがあるが、和音をベクトル化し、機能を判別しづらい和音を別の和音の加算により表現することで、和音の機能がより明確になる可能性がある。将来的に、自動作曲における代理和音などの探索にも貢献すると考えられる。

キーワード：単語埋め込み、和音進行、GloVe

Chord Progression Analysis by Word Embeddings

KOTA TSUKAMOTO^{†1} ERIKO AIBA^{†1,2} YASUHIRO MINAMI^{†1,3}

Keywords: Word Embedding, Chord Progression, GloVe

1. はじめに

1.1 機能と和声

現在、頻繁に耳にする楽曲の多くは古典的な西洋音楽の規則に基づいて作曲されたものである。このような楽曲は、様々な規則に基づいて作曲されており、その中でも機能と和声と呼ばれる考え方が多くの音楽で採用されている。機能と和声とは、個々の和音は「音階各音度上の和音の機能の観点から説明」できる[1]とする考え方であり、和音進行における規則や禁則は主として機能と和声の考えのもとにまとめられている。一方でその規則は絶対的なものではなく、実際の楽曲には規則から外れる和音進行も多く存在する。

例えば、Table.1 に示す通り、各和音にはトニック、ドミナント、サブドミナントという機能がある。トニックは安定した和音であり、主に曲の最初や最後に使われることが多い。ドミナントは不安定な和音であり、トニックに進みやすい性質を持っている。サブドミナントはやや不安定な和音であり、トニックに進むか、ドミナントを経由してトニックに進むことが多い。基本的にはトニック→ドミナント→トニック、トニック→サブドミナント→ドミナント→トニック、トニック→サブドミナント→トニックという3種類のカデンツ（和音進行のパターン）で成り立っている。このカデンツ型に従った場合、ドミナントからサブドミナントへ遷移する進行はカデンツに含まれないが、隠れマルコフモデルを用いて和音の遷移確率を学習させた研究では、ドミナントからサブ

Table.1：それぞれの調の和音と機能の対応（T=トニック，S=サブドミナント，D=ドミナント）

度数	I	II	III	IV	V	VI	VII
ハ長調	C	Dm	Em	F	G	Am	Bm-5
二長調	D	Em	F#m	G	A	Bm	C#m-5
ホ長調	E	F#m	G#m	A	B	C#m	D#m-5
へ長調	F	Gm	Am	Bb	C	Dm	Em-5
ト長調	G	Am	Bm	C	D	Em	F#m-5
イ長調	A	Bm	C#m	D	E	F#m	G#m-5
ロ長調	B	C#m	D#m	E	F#	G#m	A#m-5
機能	T	S	T(D)	S	D	T	D

ドミナントへ遷移する確率はおよそ 14%としている (Tsushima, Nakamura, Itoyama, Yoshii, 2018, 17)[2]。従って、機能と和声に従った分析を行うだけでは、不十分である。

1.2 音楽と言語

歌と日常言語は進化の過程において同じルーツを持っていると考えられている。例えば、自然言語は言語によって音の強弱、高低による方法、長短による方法などでアクセントを作っているが、これらは音楽の概念である[3]。このように、音楽と言語は共通した概念を持っている。実際、和音進行の分析には、[2]の他、隠れマルコフモデルによって得られた和音進行は全体的に一貫性がない（ある和音はその周辺の和音の影響しか受けない）

^{†1} 電気通信大学
University of Electro-Communications

^{†2} 電気通信大学技能情報学研究ステーション
Center for Art and Performance Science, University of Electro-Communications

^{†3} 電気通信大学人工知能先端研究センター
Artificial Intelligence eXploration Research Center, University of Electro-Communications

とした研究[4]など、隠れマルコフモデルや確率文脈自由文法モデルなど、自然言語処理に使われるモデルが多く使われている。

そこで、本研究においては自然言語処理モデルの一つである GloVe を用いた和音進行のベクトル化、和音ベクトルを用いた代理和音の提案などにより和音進行の規則をより明確なものとする。

本研究は、単純に統計を取るだけではなく、和音のベクトル化を行うことで和音同士の加算や減算を可能にしている。ある和音を別の和音の加算・減算で表現することで、和音の機能を細かく分類することができると考えられる。

1.3 GloVe

GloVe[5]は Pennington, Socher, Manning によって提案された単語の分散表現の獲得方法である。

GloVe は、グローバルな matrix factorization とローカルなコンテキストウインドウを組み合わせることで、カウントモデルと予測モデル、それぞれの良いところを活かして単語の分散表現を獲得できるのではないかという考えを元に提案された。

実際の実験でも、ほぼすべての条件において、今までに提案された Skip-gram[6]や CBOW[7], SVD (特異値分解) といった手法よりも意味推測 (“Athens is to Greece as Berlin is to _?”の_部分を推測する) や、構文推測 (“dance is to dancing as fly is to _?”の_部分を推測する) の点において高い精度を示している[5]。さらに、学習も高速であり、短時間でより精度の高い単語の分散表現を獲得することができる。

2. 実験

2.1 学習データ

和音をベクトル化するためのデータセットとして、ProSeMus の 9GDB(9 genres database)[8]を使用した。本データセットはバロック音楽やクラシック音楽、ロマン音楽など、9 個のジャンルごとに和音進行のデータが CHORDS ファイルとして含まれており、txt ファイルに変換可能である。和音進行は、和音同士が半角スペースで区切られて表記されている (例: C F G C)。通常の和音進行表記の他に、和音を度数 (ローマ数字) で表記したデータ (例: I IV V I) も含まれている。さらに、4 和音以上の和音を用いた和音進行のデータの他に、データ内のすべての和音を 3 和音に省略したデータも含まれている。

本研究においては、9GDB に含まれる全曲 (856 曲) の和音進行を対象にしており、和音の機能を統一するために、和音をローマ数字に置き換えたデータを用い、4 和音以上の和音を用いた和音進行のデータを用いた。

全曲の和音進行を一つの txt ファイルにまとめるにあたって、曲の最初の和音の前に”Start”という文字を、曲の最後の和音の前に”End”という文字を追加した。

データサイズは 172KB、和音の総数は 43862 個、和音の種類は 165 種類であった。

2.2 実験条件

分析手段として、統計ソフト R[9]の text2vec パッケージ[10]を使用した。

パラメータは、次元数 30、ウインドウサイズ 1、最小単語出現数 0、学習の繰り返し回数 10 とし、その他のパラメータはデフォルトのままにした。和音の種類は、自然言語処理における単語の種類より大幅に少ないため、次元数は少なめに設定した。和音は直前の和音と直後の和音による影響が大きいため、ウインドウサイズは 1 とした。最小単語出現数を 0 としたのは、和音の種類が 165 種類と、特別多くないためである。本研究で用いるデータセットはサイズが小さいため、学習の繰り返し回数は 10 回とした。

また、GloVe の分析結果との比較として、和音の共起回数の測定を行ったが、共起範囲は前後 1 和音ずつ (ウインドウサイズ 1) とした。

2.3 実験方法

2.3.1 “I” の和音に対してコサイン類似度の高い和音

“I”の和音に対する他の和音のコサイン類似度を算出し、“I”と同じ機能を持った和音 (“IVm”など) のコサイン類似度が高くなるかを調べ、コサイン類似度の高い和音を代理和音として使用できるかを考察し、本手法の妥当性を評価した。

コサイン類似度とは、2 つのベクトルがどれくらい同じ角度を向いているかを表しており、以下の式で表される。

$$\frac{\sum_{i=1}^N a_i b_i}{\sqrt{\sum_{i=1}^N a_i^2} \cdot \sqrt{\sum_{i=1}^N b_i^2}}$$

ここで、 a は和音 a のベクトル、 b は和音 b のベクトル、 N は和音の次元数を表す。

代理和音とは、主要三和音 (“I”, “IV”, “V”) と置き換えることのできる和音のことを言う。一般的には “I” の代理和音は “IIIIm”, “IVm”, “IV” の代理和音は “IIIm”, “V” の代理和音は “IIIIm”, “VIIIm-5” とされている。本実験により、これら以外の代理和音を得られると考えられる。

2.3.2 “I” の和音に対するコサイン類似度と “I” の和音に対する共起回数の比較

GloVe が和音進行に対してどのように機能するかを確かめるために、“I”の和音に対する他の和音の類似度と、“I”の和音と共起する和音の共起回数を比較した。

2.3.3 “III_m” の和音に対してコサイン類似度の高い和音

“III_m”の和音はトニックかドミナントか機能の判別が難しいため，“III_m”の和音に対してコサイン類似度の高い他の和音を調べることで，“III_m”の機能の判別を行った。

2.3.4.2 種類の和音の加算による”III_m” の和音の表現

“III_m”の和音を他の和音の加算で表現し，加算した和音の機能を元に”III_m”の和音の機能の判別を行った。ここで行う和音の加算とは，和音のベクトル化を行い，ベクトル化された和音（和音ベクトル）の中から2種類の和音ベクトルを加算することである。“III_m”の和音ベクトルに対し，2種類の和音を加算したベクトルのコサイン類似度を算出し，加算した2種類の和音の機能より”III_m”の和音の機能を判別した。

3.結果

3.1 “I”の和音に対してコサイン類似度の高い和音

実験により，“I”の和音に対してコサイン類似度の高い和音は以下ようになった。

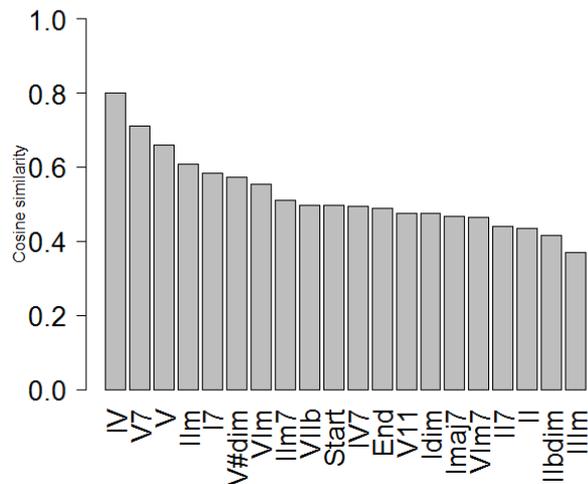


Fig.1: “I”の和音ベクトルに対してコサイン類似度の高い和音ベクトル（上位20個）

“I”の和音に対してコサイン類似度の最も高い和音は”IV”となり，その値は約0.80であった。“I”と同じトニックである”VIm”は，7番目に高いコサイン類似度であり，その値は約0.55であった。

3.2 “I”の和音に対するコサイン類似度と”I”の和音に対する共起回数の比較

実験により，“I”の和音と共起する和音の共起回数は以下ようになった。

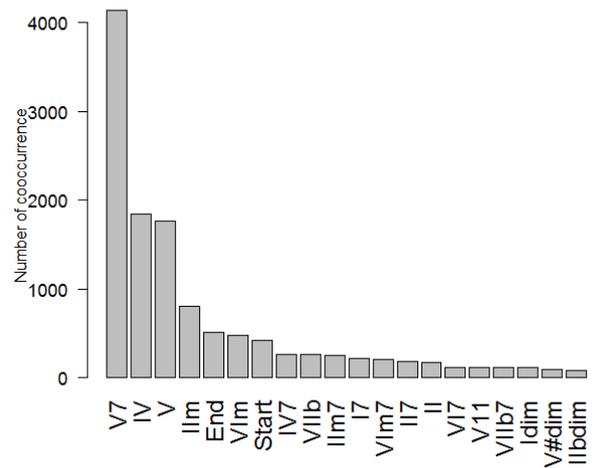


Fig.2: “I”の和音と共起する和音の共起回数(上位20個)

最も”I”の和音と共起しているのは”V7”の和音であり，共起回数は4138回であった。次に多い”IV”の和音の共起回数(1842回)に対して，約2.2倍の共起回数となった。

3.3 “III_m”の和音に対してコサイン類似度の高い和音

実験により，“III_m”の和音に対してコサイン類似度の高い和音は以下ようになった。

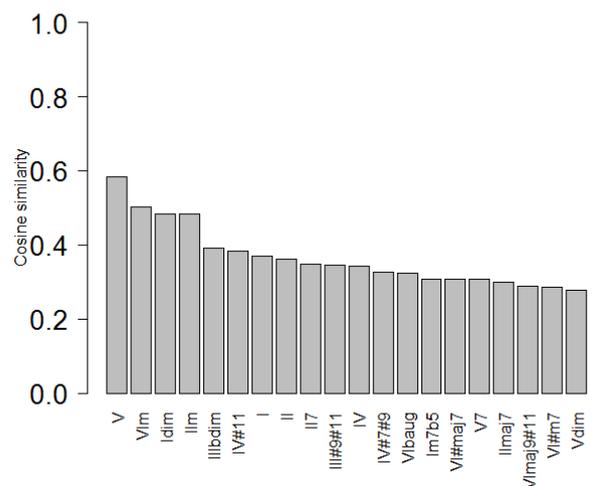


Fig.3: “III_m”の和音ベクトルに対してコサイン類似度の高い和音ベクトル（上位20個）

“III_m”の和音に対してコサイン類似度の最も高い和音は”V”であり，その値は約0.58であった。トニックの機能を持った”VIm”の和音は，2番目に高いコサイン類似度であり，その値は約0.50，“I”の和音は7番目に高いコサイン類似度で，その値は約0.37であった。

3.4 2種類の和音の加算による”III_m”の和音の表現

実験により, 2種類の和音を加算したベクトル(以降, 和音加算ベクトルと表記)と”III_m”の和音のコサイン類似度は以下ようになった。

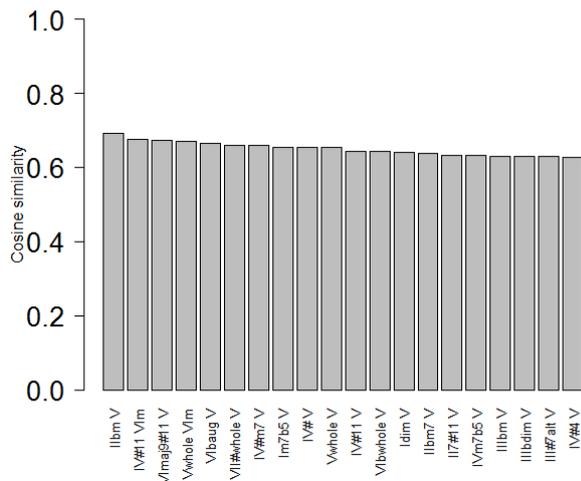


Fig.4: ”III_m”の和音ベクトルに対してコサイン類似度の高い和音加算ベクトル (上位 20 個)

本実験では, 実際の結果から”III_m”系の和音 (III_m, III_m7 など)を含んだ和音加算ベクトルは除外している。

”III_m”の和音ベクトルに対してコサイン類似度の最も高い和音加算ベクトルは”IIm V”であり, その値は約 0.69 であった。

4. 考察

4.1 “I”の和音に対してコサイン類似度の高い和音

本実験では, ”IV”の和音のコサイン類似度が最も高くなったが, ”IV”の和音の機能はサブドミナントであり, ”I”の和音に類似した和音とは言えない。コサイン類似度が高い他の和音を見ると, ”V7”や”V”などがある。これらの機能はドミナントであり, これらも”I”に類似した和音とは言えない。全体的な結果を見ると, 機能が似た和音のコサイン類似度よりも, 近接する和音のコサイン類似度の方が高くなっている。”I”の和音と機能が同じ和音である”VI_m”の和音のコサイン類似度は約 0.55 であり, この結果からは”I”の和音と”VI_m”の和音は類似しているとは言えない。

4.2 “I”の和音に対する類似度と”I”の和音に対する共起回数の比較

本実験において, ”I”の和音に対して最も共起回数が多い和音は”V7”であることがわかったが, GloVe の分析によって得た ”I” の和音に対する他の和音のコサイン類似度 (Fig.1) では, 最もコサイン類似度の高い和音は”IV”の和音であった。 ”V7”の和音の機能はドミナントであり, ”I”の和音の機能であるトニックとは大きく性質の異

なった機能であり, その代わりにサブドミナントの機能を持った”IV”の和音のコサイン類似度が高くなったことは妥当とも考えられる。

4.3 “III_m”の和音に対してコサイン類似度の高い和音

結果を見ると, 最もコサイン類似度が高い和音はドミナントの機能を持った”V”の和音であるため, ”III_m”の和音はトニックよりドミナントに近い機能を持っていると言えるが, 全体的にコサイン類似度は低く, 2 番目にコサイン類似度の高い和音がトニックの機能を持った”VI_m”の和音であるため, この結果からは”III_m”の和音の機能に関して判別することはできない。

4.4 2種類の和音の加算による”III_m”の和音の表現

全体的にコサイン類似度が低く, さらにコサイン類似度間で大きな差もないため, 信頼できる結果とは言えない。しかし, ”V”の和音を加算しているという傾向が強く出ており, ”III_m”に対してコサイン類似度の高い上位 20 個の和音加算ベクトルのうち 18 個の和音加算ベクトルが”V”の和音を含んでいた。この結果から, ”III_m”の和音はトニックよりもドミナントに近い機能を持っていると考えられる。

5. 結論

本研究では, GloVe による和音のベクトル化を行い, コサイン類似度を用いた代理和音の提案, コサイン類似度を用いた和音の機能の判別, 2 種類の和音の加算による他の和音の表現, およびその和音の機能の判別, 和音のコサイン類似度に対して和音の共起回数との比較を行った。

”I”の和音に対してコサイン類似度の高い和音の実験では, ”I”の和音に対して”IV”の和音や”V7”の和音のコサイン類似度が高いという結果が出たが, どちらも”I”の和音とは異なる機能を持った和音であるため, 代理和音として提案できるものではない。

”I”の和音に対するコサイン類似度と”I”の和音に対する共起回数の比較の実験では, ”I”の和音に対して”V7”の和音が最も多く共起しているという結果が出た。コサイン類似度では”IV”の和音の類似度が最も高くなっていたため, GloVe によってトニックと異なる性質を持ったドミナントの機能を持った和音のコサイン類似度は低くなることがわかった。

”III_m”の和音に対してコサイン類似度の高い和音の実験では, ”III_m”の和音に対して”V”の和音のコサイン類似度が最も高くなったが, ”V”の和音を含んだすべての和音のコサイン類似度が低く, 2 番目にコサイン類似度が高い和音は”VI_m”であったため, 機能を判別することはできなかった。

2 種類の和音の加算による”III_m”の和音の表現の実験では, ”III_m”の和音ベクトルに対して, ”II_bm”の和音ベクトルと”V”の和音ベクトルを足したベクトルのコサイン類似度が最も高くなった. さらに, コサイン類似度の高い他の和音加算ベクトルにおいても”V”の和音が頻繁に含まれており, この結果から”III_m”の和音はドミナントに近い機能を持っていると考えられる.

本研究で使用したデータセットは 172KB と, 非常にサイズが小さいため, 分析により和音の正確な分散表現が得られたとは言い難い. より精度の良い分散表現を得るためには, データセットを増加させるべきだと考える.

今回のような実験以外にも, 和音の連続 (II_m-V など) を一つの和音と捉えて分析を行うことで, その和音の連続の特徴の把握や, 曲中の和音ベクトルすべてを加算することで得られる曲ベクトルを元に類似した曲の検索なども可能になると考えている.

謝辞

本研究は, 平成30年度ダイバーシティ研究環境実現イニシアティブ (牽引型) および連携研究プロジェクト助成 JSPS 科研費 17H00749 の助成を受けて実施された.

参考文献

- [1]新編音楽中辞典. (音楽之友社, 2002).
- [2]Tsushima, H., Nakamura, E., Itoyama, K., & Yoshii, K. (2018). Generative statistical models with self-emergent grammar of chord sequences. *Journal of New Music Research*, 1-23.
- [3]東条敏・平田圭二(2017) 「音楽・数学・言語: 情報科学が拓く音楽の地平」近代科学社
- [4]Paiement, J. F., Eck, D., & Bengio, S. (2005). A probabilistic model for chord progressions. In *Proceedings of the Sixth International Conference on Music Information Retrieval (ISMIR)*(No. EPFL-CONF-83178).
- [5]Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).
- [6]Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).
- [7]Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- [8]Pattern Recognition and Artificial Intelligence Group - University of Alicante. “9GDB(9 genres database)”.ProSeMus. <https://grfia.dlsi.ua.es/cm/projects/prosemus/database.php>. (参照 2019-01-28)
- [9]Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. *J. Comp. Graph. Stat.* 5:299-314. Available via <http://www.R-project.org>.
- [10]Dmitry Selivanov.” text2vec”. text2vec. 2018-09-16. <http://text2vec.org/index.html> (参照 2019-01-28)