



コード進行による ヒット曲予測

+



アウトライン

01

研究動機
目的

02

研究手法

03

第一分析

04

第二分析

05

第三分析

06

研究結果

07

考察・展望

+

2年研究



01

研究動機・目的

+

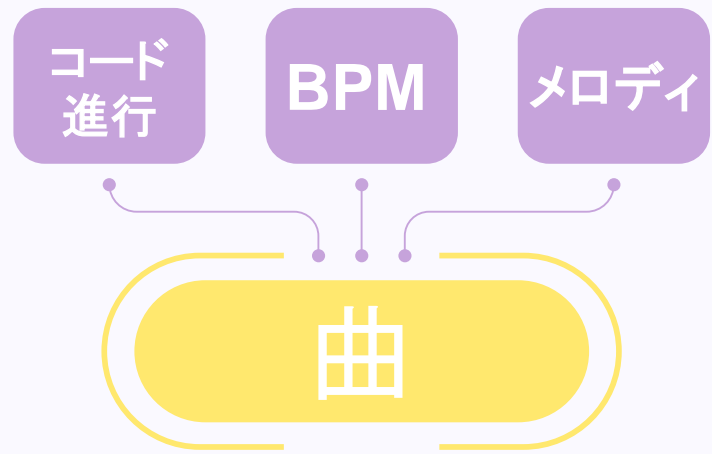


研究動機

曲はさまざまな要素を持っている



ヒット曲に法則がある



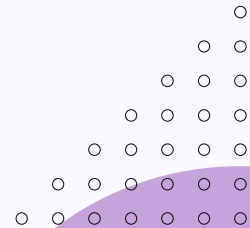
着目したのは

コード進行



目的

コード進行だけでヒット曲
を予測するAIを作成する





+

02

研究手法

....

使用データ/分析手法

+



使用データ

データを取る曲

Billboard Japan year end Hot 100

邦楽のみ

日本のアーティスト

または日本語の曲

除外

お笑い芸人の歌ネタ

歌詞の無い曲



使用データ

分析: 2010～19年各該当20曲 計200曲

式作成: 2020年各該当20曲

サビのコード進行のデータを採取



分析手法

第一分析
主成分分析

第二分析

N-gram

tf-idf

決定木分析

第三分析

LDA

ヒット曲の期待度を算出する
式を重回帰分析により作成

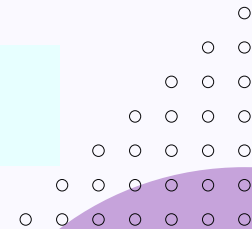


分析手法

第一分析 曲の複雑性を数値化

第二分析 曲を特徴づけるコードの組を分析

第三分析 曲をコード進行の特徴で分類



+

2年研究



03

第一分析

....

主成分分析

+



目的

コード進行の複雑性を3つの指標で評価

コード種類数

非繰り返し率

非ダイアトニックコード率

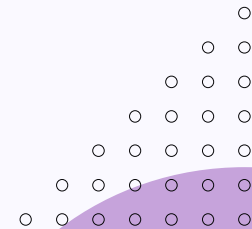
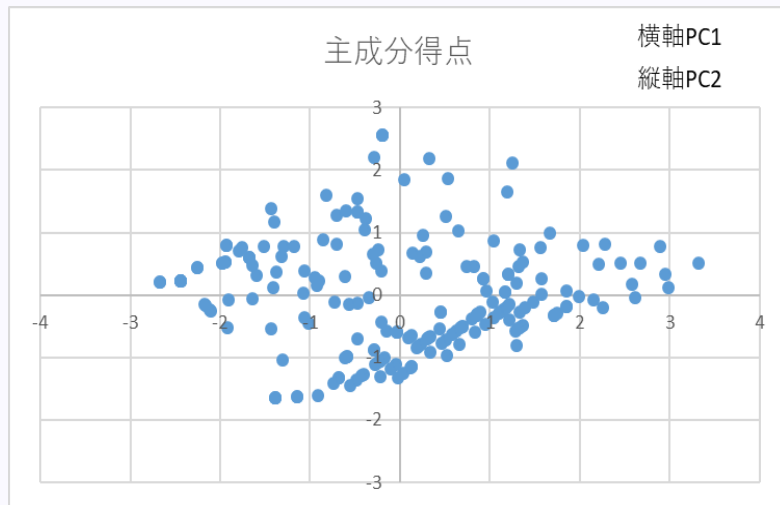
3つの指標で主成分分析

主成分得点でクラスタリング



結果

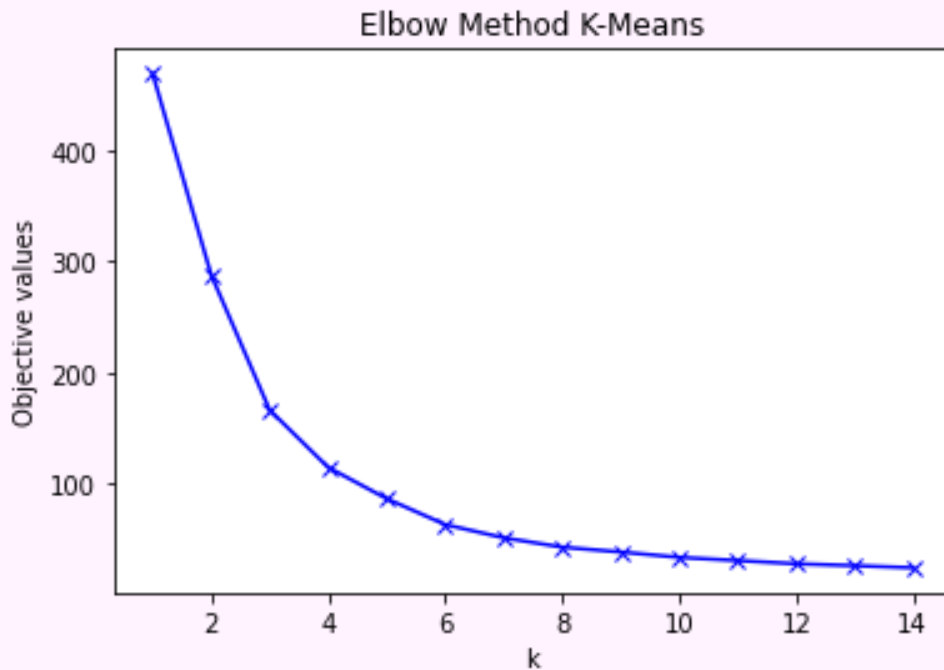
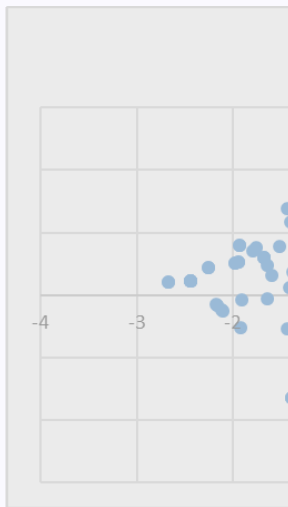
主成分得点のグラフ



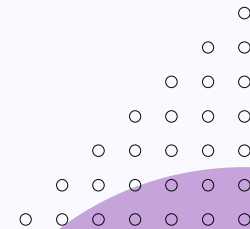


結果

主成分



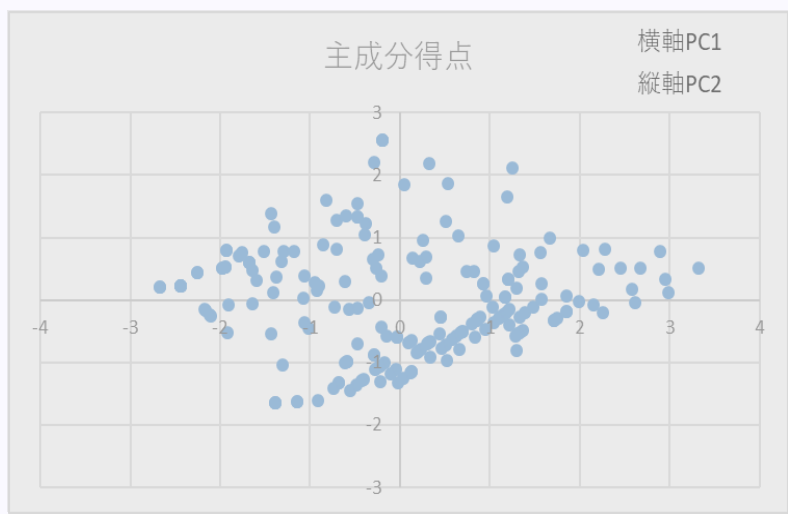
エルボー法によりクラスター数を4に決定



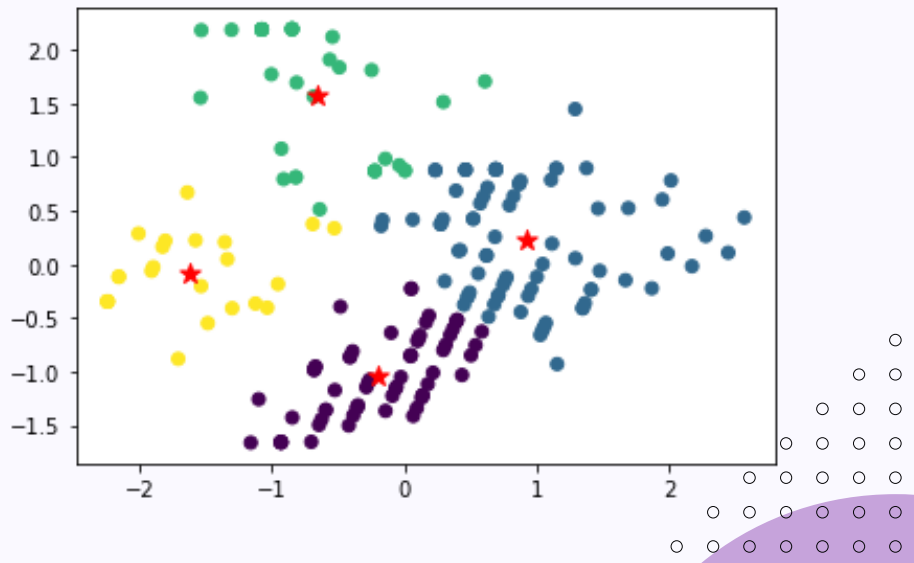
結果



主成分得点のグラフ



クラスター分析後

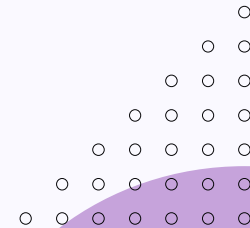




結果

作成した式

(所属したクラスタのデータ数) ÷ (代表点との距離)
をAとする



04

+

第二分析

....

N-gram/tf-idf/決定木分析

+



N-gram・・・任意の文字列を連続したn個の文字、単語で分割するデータ加工手法

例) tri-gram

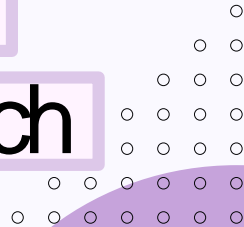
I like her songs very much

I like her

like her songs

her songs very

songs very much

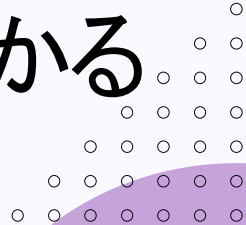




tf-idf・・・tf値(各文書におけるその単語の出現頻度)
×idf値(各単語の出現頻度の低さ)
本研究ではTri-gram後のデータを使用

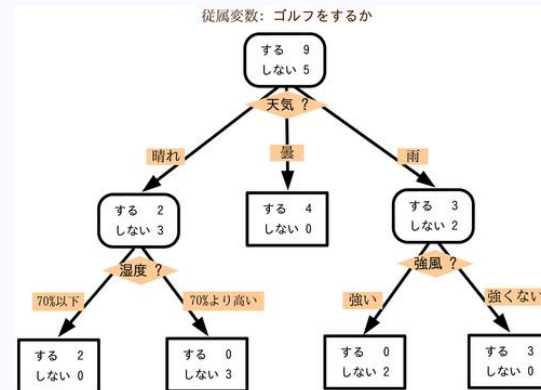


曲を特徴づけるコードの組み合わせがわかる





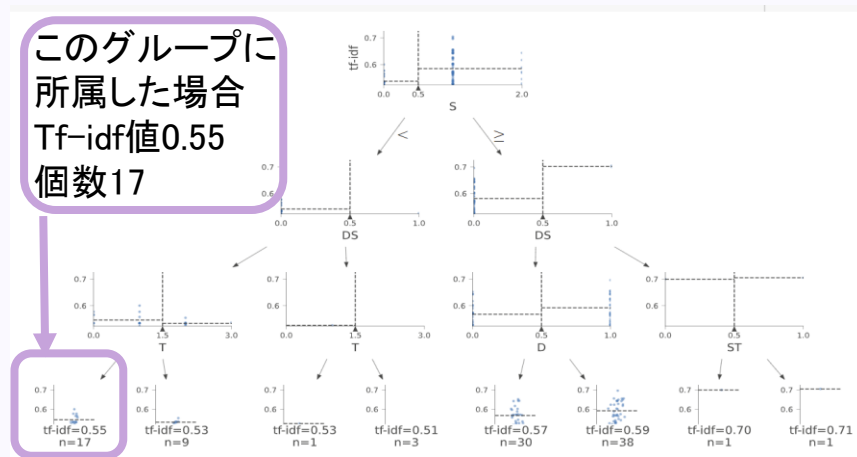
決定木分析・・・段階的にデータを分割していき 木のような分析結果を出力する 分析手法





結果

決定木分析の図



その曲tf-idf値：1曲あたり（コード数-2(個)）
現れるため、特徴を表す大きい値から3つを
採用する

（tf-idf値）+（個数）では、個数の影響が大き
いため標準化

このグループに所属した場合

（その曲のtf-idf値）² × (tf-idf値0.55 + 個数17 + 2) = **B₁**

一曲につきB₁が3つ現れるのでその積を**B**

+

2年研究



05

第三分析

....

LDA

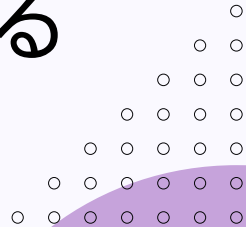
+



LDA・・・文章中の単語から文章の潜在的な
意味を確率的に推定する分析手法



曲のコード進行から曲を特徴ごとに分類できる





目的

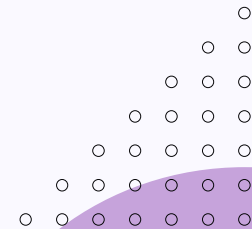
コード進行を自然言語処理



特徴をデータ化

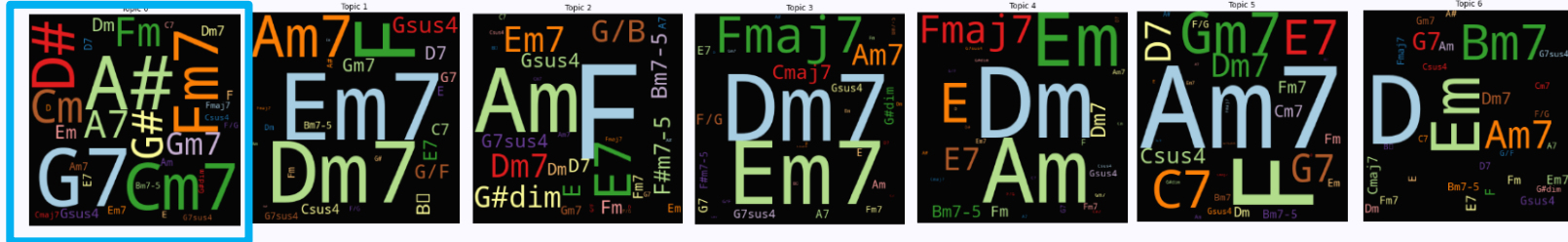


グループ分け



結果

トピック数は評価指標であるPerplexityとCoherenceから決定
各曲をコード進行から7つのトピックに分類



トピック0

トピック0: G7が0.134, A#が0.113, Cm7が0.112, D#が0.093の割合で出現する

$$\sum (k=0 \rightarrow 6) P_k \times S_k = C$$

(P_k : 演算する曲がトピック k に所属する確率)

(S_k : 分析データがトピック k に所属する確率の総和)

+

2年研究



06

研究結果

....

+



研究結果

2020版のランキング上位20曲のサビのコードを用いて、各分析結果からA・B・Cの値を算出

説明変数：A・B・Cの値

目的変数：各曲の順位

として、重回帰分析法により回帰式を作成
回帰式

$$y = -0.04 \times A - 3.25 \times B - 0.15 \times C + 21.0$$

	A	B	C
P値	0,096	0.235	0.183



+

07

考察・展望

....

+

考察



優位水準10%としてもモデル改善の余地あり

特に第二分析が不十分

	A	B	C
P値	0,096	0.235	0.183

展望

2年研究



各年上位20曲

比較

それ以外の曲

システムの構築

コード進行の情報

AIがヒット曲を予測

謝辞



本研究にあたり，深山覚主任研究員（国立研究開発法人産業技術総合研究所）に指導助言いただきました。厚くお礼申し上げます。

