

単語分散表現を用いた観光地レビュー  
からのクロスドメイン歌詞検索

韓毅弘 山西良典 西原陽子（立命館大学）

奥健太（龍谷大学）

# 研究背景

- 音楽聴取方式の変化
- 聞く場所は自由，アクセスできる音楽数はほぼ無制限



# 研究目標

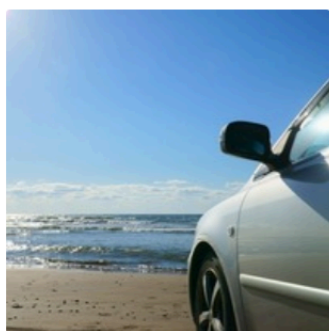
- 観光地や雰囲気に適した音楽を流す
- 実体験しているものと協調する音楽聴取が可能
- 観光活動にBGMをつける→観光体験の向上



<https://www.hemingway.cafe/>



3 ©鴨志田一/KADOKAWA アスキー・メディアワークス/青ブタ Project



## 海で聴くならこの音楽を！海のBGMに最適な曲10選♪

夏になると海にドライブやデート、海水浴などにでかける人も多いのではないのでしょうか？海でのんびりするときや、海へのドライブ中や車の中で聴くのにおすすめのBGMをご紹介します。今だから聴きたい！と思える曲を中心に厳選してみました！周りの迷惑にならないように、音楽を楽しんじゃいましょう♪

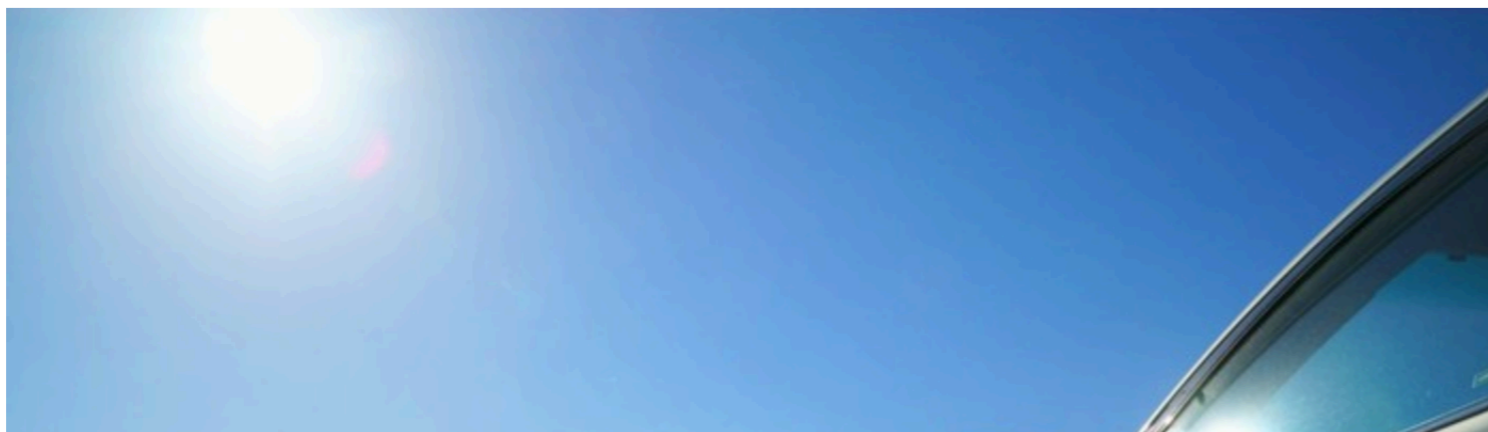
全国 #音楽 海



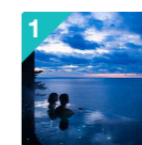
cherry



## 海までドライブ！波の音とともに、音楽を楽しもう♪



### 👑 全国のランキング(月間)



週末サクッと温泉旅行！東京から気軽に行ける温泉地6選



【閲覧注意】観光地だと思ってたのに…！戦慄の全国心霊スポット10選



ニトリのプチプラ商品がすごい！主婦厳選のおすすめ商品12選



海で聴くならこの音楽を！海のBGMに最適な曲10選♪

<https://icotto.jp/presses/10294>



オープニングテーマ

『君のせい』

the peggies

作詞・作曲:北澤ゆうほ

編曲:the peggies, 川口圭太

(EPIC Records Japan)

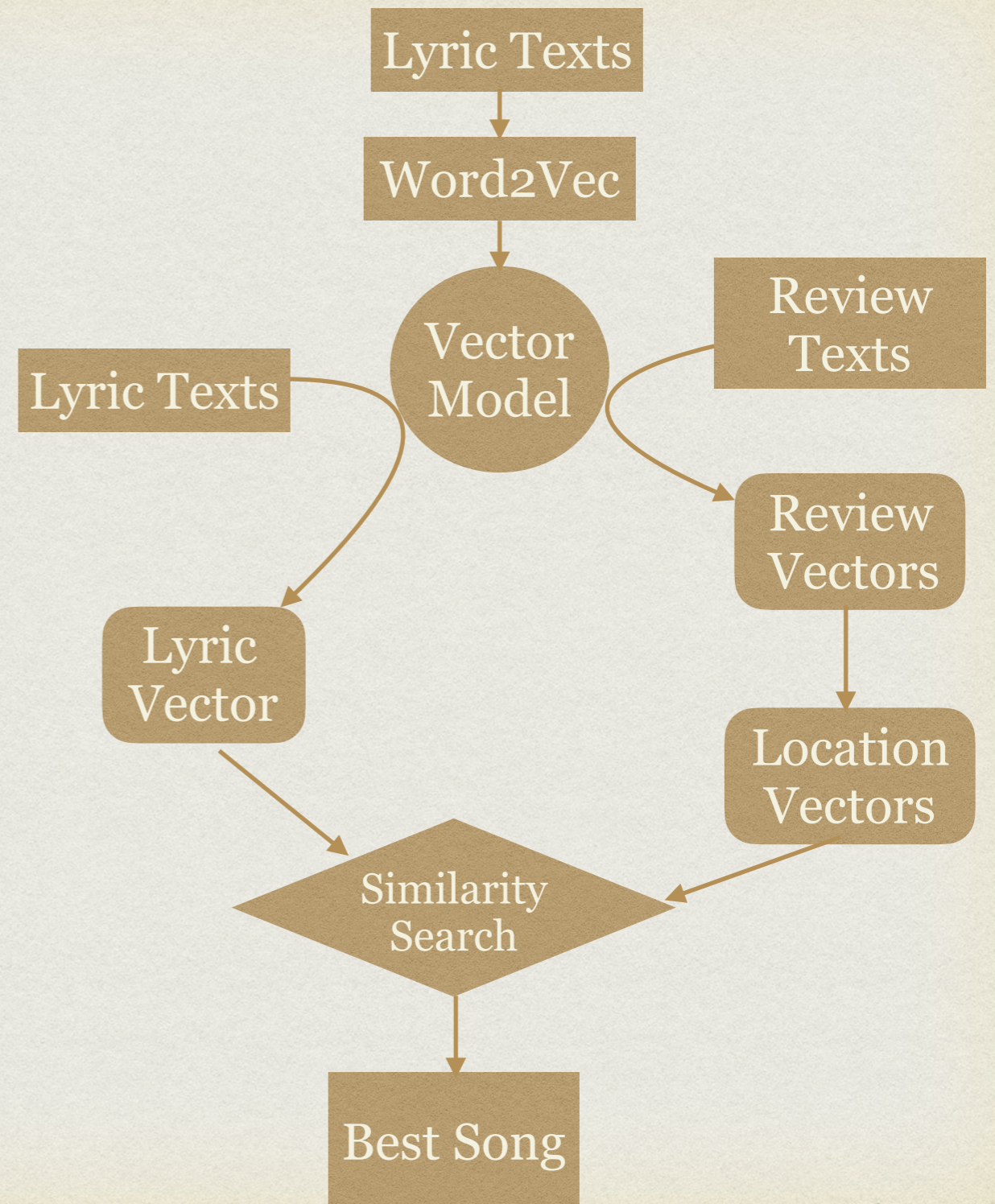
©鴨志田 一 / KADOKAWA アスキー・メディアワークス / 青ブタ Project

# コンセプト

- 観光地や雰囲気に適した楽曲の検索&推薦
- 観光地の雰囲気に対する一般的評価→観光地レビュー
- 楽曲の性質を表す媒介→歌詞テキスト
- 歌詞テキストで単語分散モデル構築
- 単語分散モデルで観光地レビューをベクトル化→レビューを歌詞として解釈
- 歌詞として解釈されたレビューに類似度が高い歌詞を検索  
⇒観光地と音楽のクロスドメイン検索

# 手法

- 歌詞テキストから単語分散モデルを構築
- 単語分散モデルで歌詞データと観光地レビューをベクトル化
- レビューベクトル群から観光地ベクトルに統合
- 観光地ベクトルと歌詞ベクトルと照合
- 類似度の高い歌詞と選出



# 単語分散表現

<https://ja.wikipedia.org/wiki/電車>



電車

台湾・高雄メトロの飲食禁止ポスター  
(C) Simon Creative

+



萌え

=

<http://caf-aholic.seesaa.net/article/224434012.html>



萌え電車

- 各概念は複数の計算要素で表現される
- 各計算要素は複数の概念の表現に関与する



# 単語分散表現モデルの構築

- 歌詞サイト azlyrics から歌詞データを取得
- 英語歌詞のみで構成された歌詞コーパスを作成
- 歌詞コーパスには全94451件の英語歌詞が含まれる
- Word2Vecで単語分散モデルを構築

Word2Vecパラメータ : size = 300, window = 10, min count = 2, workers = 8, iter = 10

# 歌詞と観光地レビューのベクトル化

- 単語分散表現モデルを参照して単語ベクトルを得る
- 各単語のベクトルを加算して単語数で割ることで歌詞やレビューの平均ベクトルを計算する
- 但し、歌詞コーパスにはない単語がレビュー文に存在する単語ベクトルが得られない場合、その分の単語数を消す

- $$\bar{V} = \frac{\vec{v}_1 + \dots + \vec{v}_n}{n}$$

# レビューベクトル群から 観光地ベクトルの統合

- 個人の主観的評価を統合→観光地の一般的性質が得られる
- レビュー中の文字数で重み付けされた平均計算手法
- $$\bar{X} = \frac{\omega_1 \vec{x}_1 + \dots + \omega_n \vec{x}_n}{\omega_1 + \dots + \omega_n}$$
- レビューの文字数の違いによる情報量貢献の差を表現
- 統合結果が観光地性質を表現する信頼性が向上

# 歌詞と観光地のベクトル 類似度による歌詞選曲

- 歌詞ベクトルと観光地ベクトルのコサイン類似度を計算

- $$\cos(\vec{X}, \vec{V}) = \frac{\vec{X} \cdot \vec{V}}{|\vec{X}| \times |\vec{V}|}$$

- 類似度の高い歌詞はこの観光地の雰囲気に近いと見なす
- 類似度が最も高い歌詞を観光地に適した歌詞として出力

# 実行結果

観光地と選曲される音楽の番号のリスト(一部抜粋)

観光地	推薦歌詞番号
Bowling Green	39663
Ellis Island Immigration Museum	84063
Conservatory Garden	55628
Riverside Park	74814
Hudson River Park	74814
Roosevelt Island	39663
Fort Troon Park	74814
New York Harbor	56837
Franklin D Roosevelt Four Freedoms Park	53520
Long Beach	54217

# Ellis Island Immigration Museum

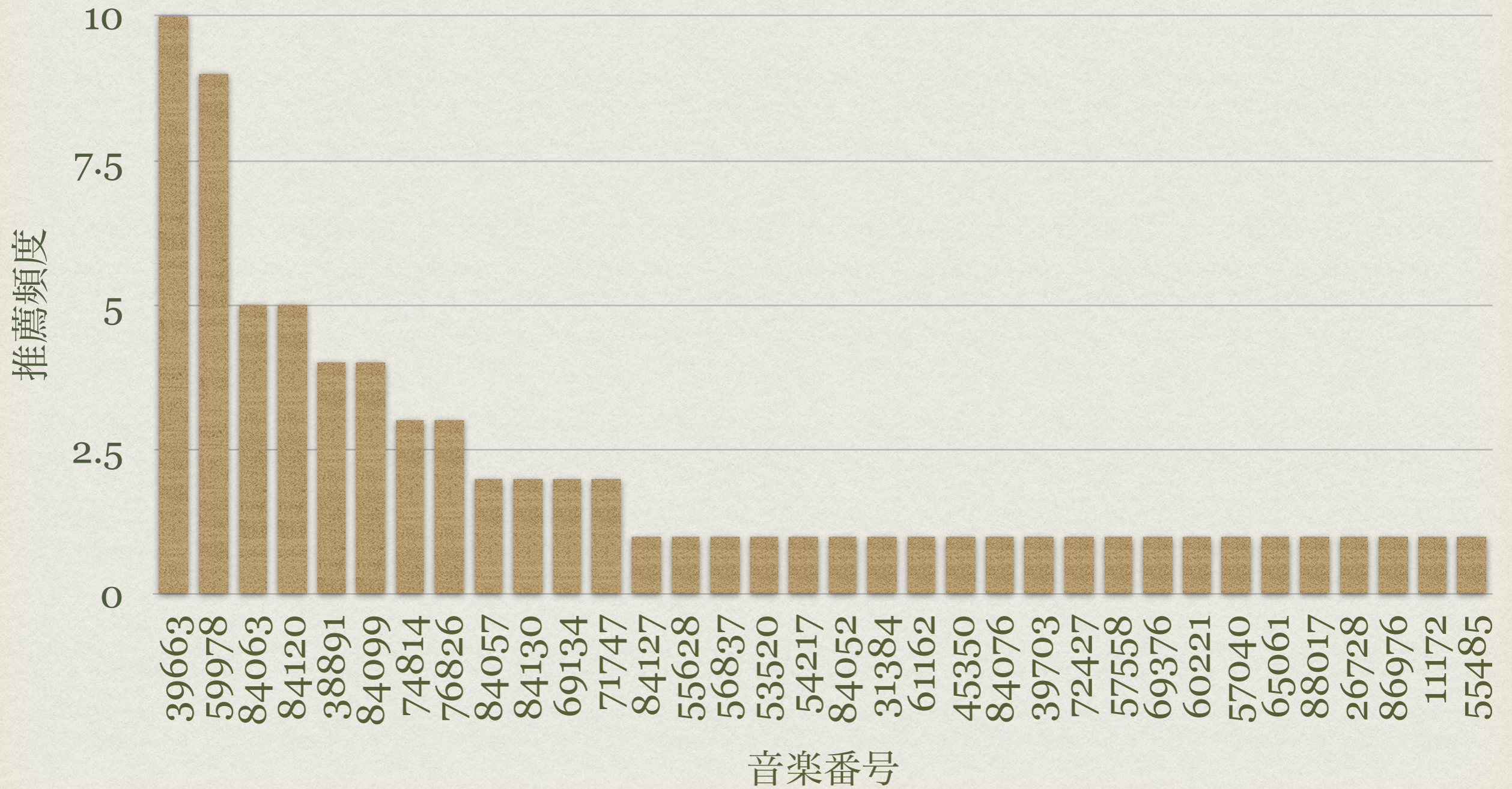


Roosevelt Island



Riverside Park

# 実行結果

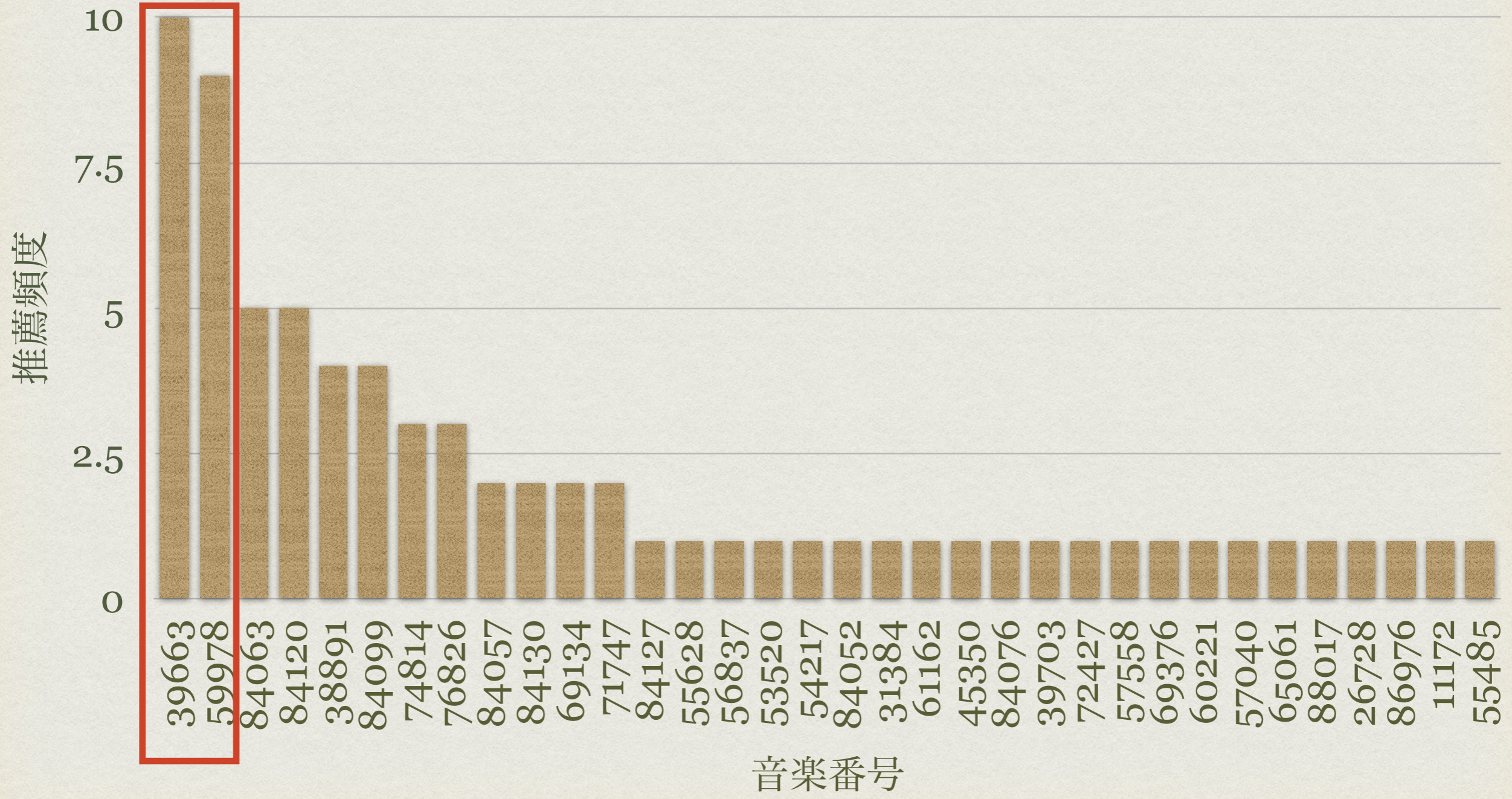


# 選曲の偏りについての考察

- 複数の観光地に対して同じ音楽が選出される選曲の偏りが確認された
- 原因は観光地に共通点があると考えられる
- 多くの観光地に適する音楽にヒットされた観光地をリストアップ
- 番号39663と59978にヒットされた観光地について考察



# 実行結果



# 選曲の偏りについての考察

観光地名	観光地所在
Roosevelt Island	New York, USA
Bowling Green	Kentucky, USA
Governors Island National Monument	New York, USA
SoHo	New York, USA
West Village	New York, USA
Meatpacking District	New York, USA
Twin Peaks	California, USA
Lincoln Park Conservatory	Illinois, USA
Greenwich	London, UK
Greenwich Park	London, UK

観光地名	観光地種類
Neue Galerie	博物館
Solomon R Guggenheim Museum	博物館
New York Historical Society Museum Library	博物館
Museum of Arts and Design	博物館
United Nations Headquarters	国連本部
Broadway	街並み
Radio City Music Hall	劇場
Le Puy du Fou	テーマパーク
Westminster	街並み

# 選曲の偏りについての考察

- 音楽番号39663にヒットされた観光地の共通点：**所在地**
- アメリカ，特にニューヨークにある観光地が多い
- Greenwich と Greenwich Park は場所的に同一視
- 音楽番号59978にヒットされた観光地の共通点：**歴史関連**
- 博物館が多い，劇場は歴史名所，テーマパークが歴史テーマ
- Westminster は歴史的な町

# まとめ

- 観光地と歌詞のクロスドメイン音楽検索手法を提案した
- 単語分散モデルで観光地レビューをベクトル化することで観光地を歌詞として解釈，歌詞との類似度計算が可能に
- 選曲結果での偏りの原因について考察した
- 選曲結果に対しての主観評価については今後の課題とする
- 他の推薦アルゴリズムとの組み合わせも検討する

ご清聴ありがとうございました