

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^{†, a} 山西良典^{†, b} 西原陽子^{†, c} 奥健太^{†, d}

† 立命館大学 † 龍谷大学

a) is0387ps@ed.ritsumei.ac.jp b) ryama@media.ritsumei.ac.jp c) nisihara@fc.ritsumei.ac.jp

d) okukenta@rins.ryukoku.ac.jp

概要 本稿では、単語分散表現を用いて観光地レビューに類似したコンテキストの歌詞を選曲する手法を提案する。聴取する音楽を選曲する上では、ジャンルやアーティストのようなメタデータ以外にユーザの聴取環境の利用が考えられる。ユーザの聴取環境を考慮して音楽を選曲可能となれば、ドライブ中の景色に適した音楽や観光中に観光スポットの雰囲気や背景に適した楽曲を視聴可能となる。本稿では、単語分散表現のモデルをドメイン間で共用するクロスドメイン検索を用いることで、観光地の雰囲気や背景に適した歌詞検索を実現する。観光地のレビューの単語ベクトル化において、歌詞コーパスを用いて構築された分散表現モデルを用いることで、観光地レビューを歌詞としてみなした単語ベクトルを得る。これを利用して観光地レビューにベクトル類似度の近い歌詞を検索する。このとき、共用の分散表現モデルの構築に異なるコーパスを用いた場合の検索結果との比較・分析を行った。

キーワード 音楽情報処理, 歌詞, コンテキストアウェア選曲, クロスドメイン検索

1 はじめに

音楽は、電子機器とウェブの発展に伴って、インタラクティブモデルが変化してきた代表的なエンタテインメントである。ユーザが音楽を聴取する形態の変化に着目してみると、従来はコンサートホールやライブハウスのように音楽に特化した場所に紐付けられて聴取されていた。現在は、携帯音楽再生機器や車中においても容易に多数の楽曲聴取が可能になり、音楽聴取の場面は様々に広がった。また、サブスクリプションサービスの登場によって世界中の音楽を自由に聴取可能となり、ユーザのライブラリに保持される楽曲数の上限はほぼ無限大である。つまり、従来は特定の音楽のみを楽しむ環境であったが、どこでも、どんな音楽でも聴けるようになり、何か他のアクティビティを行いながら音楽を聴取するというエンタテインメントの並行性が生まれるようになってきた。このような背景のもとで、ユーザの聴取コンテキストと協調的に聴取される音楽の楽しみ方（コンテキストアウェアな選曲）が考えられる。本稿では、音楽を楽しむユーザの聴取コンテキストの一つとして観光時の音楽聴取を考える。歌詞の中に特定の地名が入っている場合、その土地を旅行で訪れたタイミングでそのような楽曲を聴取した経験を有する人も多いであろう。例えば、アメリカのサンフランシスコで Grantis の “San Francisco” や日本の大阪で Coldplay の “Lovers in Japan” を聴取することで観光地と歌詞の一致は観光はより印象深くなる。また、具体的な地名が入っていな

くても、観光スポットの雰囲気や特性に合致した音楽を聴取することが可能になれば観光から得られる印象はより強いものへとなることが期待される。例えば、広く穏やかな公園である Vancouver の Stanley Park では、Ed Sheeran の “Perfect” のように穏やかで慈愛に満ちた楽曲、東京の歌舞伎町のような歓楽街では System of a Down の “Violent Pornography” などの刺激的な楽曲をそれぞれ聴取することで、観光の体験は音楽聴取によって印象深いものへと変化するであろう。これは、映画やアニメなどの映像コンテンツにおいて、各シーンに合わせた BGM が再生されることを現実の生活の中で楽しめるようにすることに等しい。本稿では、観光時の音楽聴取によって観光体験をより豊かにすることを目的とし、観光地に適した歌詞を有する楽曲の検索技術を提案する。提案手法では、観光地のレビューを観光地の雰囲気に対する一般的な評価、あるいは、経験の集合知としてみなすことで、観光地レビューをクエリとした歌詞検索（つまり、観光地と音楽のクロスドメイン検索）を実現する。このとき、歌詞を学習データとした単語分散表現モデルを構築し、この単語分散表現モデルによって観光地レビューをベクトル化することで、観光地レビューを歌詞として解釈する。そして、歌詞として解釈された観光地レビューに類似度が高い歌詞を検索することで、観光地からの楽曲検索を目指す。

2 関連研究

音楽検索研究は世界中で数多く実施されている。例えば、鼻歌によるメロディ検索 [1] や音楽ジャンル分

類 [2, 3] は、音楽情報処理に関する国際会議で頻繁に議論されているテーマである。音楽、特に歌唱曲、は音響的側面（楽音）と言語的側面（歌詞）の組み合わせによって表現される芸術であり、音楽自身が複数のメディアによって形成されるマルチメディアと言える。つまり、歌唱曲の聴取から得られる感性は「音響聴取」と「歌詞理解」の2種類の認知作用によって形成される。しかしながら、従来の音楽検索システムの多くは音響特徴量のみ注目したもの [4] が多く、音楽の言語的側面（つまり、歌詞）に着目した研究は少ない。歌詞特徴を用いた楽曲検索としては、Tsukuda らの研究 [5] や Cai らの研究 [6] が挙げられる。Tsukuda ら [5] は、Lyric Jumper というサービス上で、歌詞トピックが類似するアーティストを推薦する機能を導入している。Cai ら [6] は、ブログ等の Web 文書を閲覧中に音楽を推薦する MusicSense を提案している。Web 文書と音楽とのマッチングには、文書と歌詞それぞれから抽出された感情語が用いられる。Cai らの研究では、“title + lyrics” という形式のクエリで検索された文書を歌詞情報として用いているのに対し、本研究では事前に収集した歌詞コーパスから構築した単語分散表現モデルを用いている点で異なる。

音楽と位置情報のマッチングについては、これまでも研究されている。Kaminskas ら [7] は、位置情報ウェア楽曲推薦システムとして、タグベースアプローチ [8] および知識ベースアプローチ [9] を提案している。タグベースアプローチ [8] では、音楽および POI（関心地点）に対し付与されたタグを基に音楽と POI を関連付ける。知識ベースアプローチ [9] では、DBPedia を知識ベースとして用いて音楽と POI を意味的に関連付けたグラフを構築し、与えられた POI に関連する楽曲をランキングする。さらに、これらタグベースおよび知識ベースアプローチを組み合わせたハイブリッドアプローチ [7] も提案されている。Kaminskas ら [7] の分類に対し、本研究は歌詞ベースアプローチに位置付けられる。

一方で、言語処理に関する研究では、インターネットの普及に伴って校正された自然言語ではなく一般ユーザが記述した多様な一見不自然とも見える言語を対象とした研究 [10, 11] は 2010 年前後から盛り上がりを見せ、近年ますます注目されている [12, 13]。しかしながら、小説や歌詞といった芸術・文化の一部として捉えられる文書に対しての研究は少ない。本研究は、音楽の言語的側面に着目し、音楽の検索において人間のアクティビティを入力とするインタラクションモデルの提案と位置付けられる。

3 提案手法

本稿では、単語分散表現 [14] を用いて観光地レビューに類似したコンテキストの歌詞を選曲する手法を提案する。提案手法では、英語歌詞のコーパスを用いて単語分散表現モデルを構築する。構築した単語分散表現モデルによって歌詞と観光地に付与された各レビューそれぞれの単語分散表現ベクトルを獲得する。ここで、任意の観光地に付与されている全てのレビューの平均ベクトルを得ることで、観光地単位での単語分散表現ベクトルを得る。観光地と歌詞の単語分散表現ベクトル間の類似度を計算し、観光地と類似度が高いベクトルをもつ歌詞を検索結果として出力する。提案手法のコンセプトは、観光地レビューのテキストを一種の歌詞として見なし、レビューを「歌詞」という次元に統一させることで観光地と音楽の類似度を類似性を計算可能にし、観光地に対する音楽を選曲することである。提案手法の全体の流れは以下のようなになる (図 1 参照)。

処理 1 歌詞コーパスを用いた単語分散表現モデルの構築

処理 2 歌詞と観光地レビューの単語分散表現ベクトル化

処理 3 レビュー単位から観光地単位の単語分散表現ベクトルへの統合

処理 4 歌詞と観光地の単語分散表現ベクトル類似度に基づく歌詞検索

以下では、それぞれの処理について詳細を説明する。

3.1 単語分散表現モデルの構築

歌詞サイト azlyrics¹ から取得した歌詞データを用いて単語分散表現モデルを構築する。ここで、ひらがなや漢字といったアルファベット以外の文字種が用いられている歌詞、英単語で記述されていない歌詞（例えば、ローマ字表記で日本語の発音が表現されている歌詞）を削除対象とする。すべての歌詞から英語以外の歌詞を取り除き、英語歌詞のみで構成された歌詞コーパスを作成する。歌詞コーパスには、全 94,451 件の英語歌詞が含まれる。

このコーパスを Word2Vec (Skip-Gram) [15, 16] の訓練用データとして入力し、各単語を 300 次元のベクトルとして表現する単語分散表現モデルを構築する。このとき、Skip-Gram の設定としては、 $size = 300$, $window = 10$, $min_count = 2$, $workers = 8$, $iter = 10$ とした。

3.2 歌詞と観光地レビューのベクトル化

英語歌詞データから構築された単語分散表現モデルを用いて、歌詞と観光地レビューの単語分散表現のベクトルを獲得する。具体的には、歌詞や観光地レビューの

¹<https://www.azlyrics.com/>

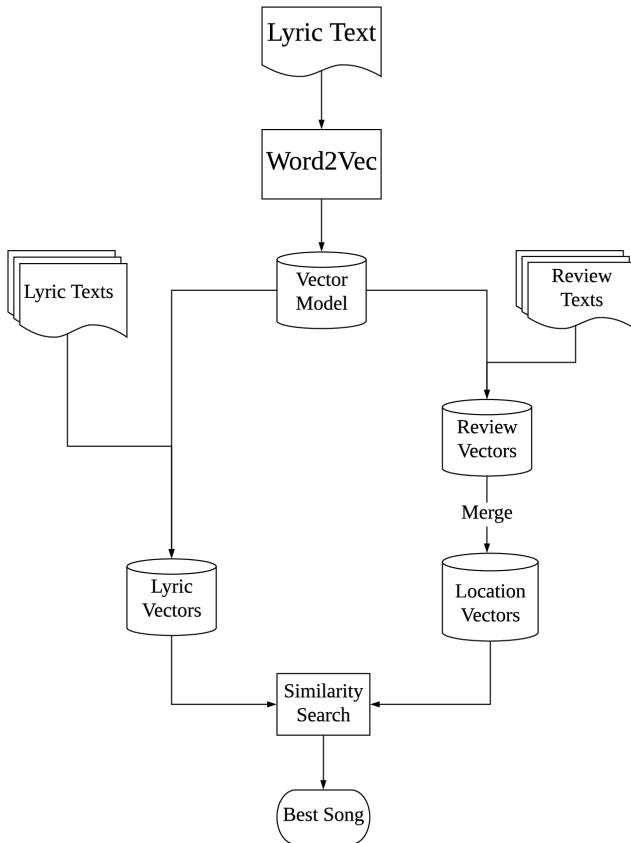


図1 システムのフローチャート

各単語に対して、単語分散表現モデルを参照して単語ベクトルを得る。歌詞や観光地レビュー文に含まれる各単語のベクトルを加算し、レビュー中に出現する単語数で除算することでレビューあたりの平均ベクトルを計算する。

ただし、単語分散表現モデルは歌詞コーパスを用いて訓練したものであるため、観光地レビューで出現する単語（例えば、具体的な地名など）の分散表現ベクトルがモデルに含まれていないケースがある。このようなモデルに含まれない単語を含んだレビューの平均ベクトルを計算する時の処理として、上述の平均ベクトル取得の際に当該単語を加算および除算の対象から取り除く。歌詞および各レビューのベクトル化は以下の式 (1) に従って各単語の相加重で計算する。

$$\bar{\mathbf{V}} = \frac{\vec{v}_1 + \dots + \vec{v}_n}{n}, \quad (1)$$

ここで、 \vec{v} は各単語のベクトル、 $\bar{\mathbf{V}}$ は歌詞の平均ベクトル、 n は単語数である。

3.3 観光地レビューベクトルから観光地ベクトルの作成統合

観光者によって記述された観光地レビューには、レビュー個人感情や主観的な評価が含まれている。同一の観光地に付与されたレビューを観光地単位に統合することで、観光地の個性や一般的な性質を表現可能になると考える。ベクトル統合の計算法については、平均の計算法の違いにより、以下の2種類を用意した。

1. 相加重平均手法 各レビューのベクトルの加算和をベクトル数で除算した結果を観光地ベクトルとする。

$$\bar{\mathbf{X}} = \frac{\vec{x}_1 + \dots + \vec{x}_n}{n}. \quad (2)$$

ここで、 \vec{x} は各レビューのベクトル、 $\bar{\mathbf{X}}$ は観光地の平均ベクトル、 n は観光地のレビュー数である。

2. 加重平均手法 各レビューのベクトルについて、レビュー中の文字数で重み付けされた平均値を観光地ベクトルとする。

$$\bar{\mathbf{X}} = \frac{\omega_1 \vec{x}_1 + \dots + \omega_n \vec{x}_n}{\omega_1 + \dots + \omega_n}. \quad (3)$$

ここで、 \vec{x} は各レビューのベクトル、 ω は各レビューの文字数、 $\bar{\mathbf{X}}$ は観光地の平均ベクトル、 n は観光地のレビュー数である。

相加重平均手法ではレビューの文字数を考慮しないため、レビュー長に大きな差がある場合、文字数の少ないレビューと多いレビューで観光地ベクトルに与える影響の重みが等しくなる。その結果、観光地のレビュー中に一言で記述された簡素なレビューが多く現れる場合、観光地ベクトルとしての信頼性が低くなると考えられる。一方で、加重平均手法はレビューの文字数によって、各レビューの観光地に対する性質表現における貢献度を評価する。これは、文字数の多いレビューは観光地に関する情報量も多いという仮説に基づく。

3.4 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類似度を計算し、観光地ベクトルとの類似度の高い歌詞はこの観光地の雰囲気に近いと見なして楽曲を選出する。具体的には、各観光地ベクトルに対して全ての歌詞ベクトルとのコサイン類似度を計算する。そして、類似度が最も高い歌詞を入力された観光地に適した歌詞として出力する。

$$\cos(\vec{q}, \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{|\vec{q}| \times |\vec{d}|}. \quad (4)$$

ここで、 \vec{q} と \vec{d} はベクトル、 $\cos(\vec{q}, \vec{d})$ は \vec{q} と \vec{d} のコサイン類似度である。

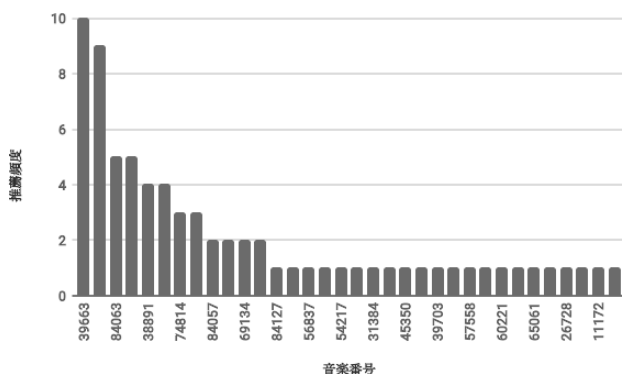


図2 歌詞に対して選曲された観光地数.

4 実験

評価実験では、提案手法のコンセプトである「レビューを歌詞に見なす」ことの有効性を検証するため、歌詞テキストではなく、レビューテキストで構築した単語分散表現モデルを利用した比較システムを作成した。比較手法の構築は以下のように示す。

手順1 観光地レビューのコーパスを用いた単語分散表現モデルの構築

手順2 歌詞と観光地レビューの単語分散表現ベクトル化

手順3 レビュー単位から観光地単位の単語分散表現ベクトルへの統合

手順4 歌詞と観光地の単語分散表現ベクトル類似度に基づく歌詞検索

表1に、提案手法と比較手法の選曲結果（一部抜粋）をそれぞれ示す。ただし、音楽番号と歌詞URLのリスト（一部抜粋）は表2に示す。図2に、選曲される各歌詞が選出された観光地数を表す。これらの結果から、以下のことが確認された。

1. 提案手法による選曲結果と比較手法による選曲結果は大きく異なる。
2. 選曲結果では、特定の楽曲が複数の観光地に集中して選曲されている。
3. Riverside ParkとHudson River Parkは、どちらの手法でも同一曲が選ばれている。また、他にも「Park」が語尾につく名称の観光地には同一の楽曲が選ばれた傾向にある。

5 考察

実験結果について、選曲の偏りとデータソースの2種類の観点から考察する。

5.1 選曲の偏りについての考察

選曲結果では、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。この原因として、観光地についての共通点が考えられる。この仮説を検証するため、最も多くの観光地に対して選曲された音楽番号39663と59978について、選出対象となった観光地をそれぞれ表3と表4にリストアップした。

その結果、いずれの観光地にも共通点が発見された。音楽番号39663が選曲された観光地はアメリカ、特にニューヨーク州にあることが多いことがわかった。GreenwichとGreenwich Parkについては、そもそも後者は前者の中に含まれる公園であるため、同一視可能であり、同一の選曲がなされたとしても問題がないと思われる。すなわち、音楽番号39663が選曲された観光地は、位置的な共通点があると考えられることは妥当であると思われる。音楽番号59978が選曲された観光地は、博物館が多く、他にも歴史関連のテーマパークや歴史のある劇場がある。WestminsterはBroadwayと同じく「街並み」に関する観光地ではあるが、Broadwayとは異なり、歴史名所の多い町である。結果として、音楽番号59978が選曲された観光地は、共通して歴史に関する観光地であることが多かった。

文献[17]においても指摘されているように、位置情報と音楽とのマッチングにおいて選曲結果を多様化することについては、この研究トピックの共通の課題として挙げられる。今後は、選曲結果を上位1件に限定せず、多様化する工夫についても考えていく。

5.2 データソースに対する考察

提案手法では、歌詞という感性的なメディアを選出するため、観光地に関する情報として人間の主観的意見を含む観光地レビューを用いた。一方で、観光客による観光地レビューは、個人の経験や旅行中の出来事など観光地自体の情報ではない記述や単語も散見された。そのため、観光地レビューは観光地自体との関連性が低い可能性もある。

この問題を解決するためには、観光地レビュー以外にも観光地の公式情報などの紹介文を利用する可能性が考えられる。観光地の紹介文を使うことで、観光地自体と情報の関連性は保証されることになる。一方で、欠点として、感情表現などが少なくなり、感性的なメディアである歌詞の検索には不十分となる恐れもある。

6 おわりに

本稿では、単語分散表現を用いて観光地レビューに類似したコンテキストの歌詞を選曲するクロスドメイン音楽検索手法を提案した。歌詞コーパスを学習することによって構築された単語分散表現モデルに基づいて、観光

表 1 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	提案手法	比較手法
The Montcalm at the Brewery London City	84127	92790
The Beekman A Thompson Hotel	84057	80687
Conservatory Garden	55628	35313
Riverside Park	74814	56035
Hudson River Park	74814	56035
Roosevelt Island	39663	20655
Fort Troon Park	74814	20655
New York Harbor	56837	39661
Franklin D Roosevelt Four Freedoms Park	53520	20655
Long Beach	54217	54215

表 2 音楽番号と歌詞 URL のリスト (一部抜粋)

音楽番号	歌詞 URL
84127	https://www.azlyrics.com/lyrics/sunkilmoon/strangerthanparadise.html
92790	https://www.azlyrics.com/lyrics/whitestripes/littlecreamsoda.html
84057	https://www.azlyrics.com/lyrics/sunkilmoon/beautifulyou.html
80687	https://www.azlyrics.com/lyrics/slimdusty/themanfromthenevernever.html
55628	https://www.azlyrics.com/lyrics/lobo/whyisitme.html
35313	https://www.azlyrics.com/lyrics/gregoryalanisakov/fireescape.html
74814	https://www.azlyrics.com/lyrics/rodstewart/manhattan.html
56035	https://www.azlyrics.com/lyrics/lorettalynn/imshootinfartomorrow.html
39663	https://www.azlyrics.com/lyrics/idinamenzel/oneshortday.html
20655	https://www.azlyrics.com/lyrics/cowboyjunkies/arlington.html
56837	https://www.azlyrics.com/lyrics/luckyboysconfusion/likeratfromasinkingship.html
39661	https://www.azlyrics.com/lyrics/idinamenzel/nothininthisworld.html
53520	https://www.azlyrics.com/lyrics/lilb/walktheworld.html
54217	https://www.azlyrics.com/lyrics/limahl/tarbeach.html
54215	https://www.azlyrics.com/lyrics/limahl/staywithme.html

表 3 音楽番号 39663 が選曲された観光地と観光地の所在.

観光地名	観光地所在
Roosevelt Island	アメリカ・ニューヨーク州
Bowling Green	アメリカ・ケンタッキー州
Governors Island National Monument	アメリカ・ニューヨーク州
Meatpacking District	アメリカ・ニューヨーク州
Lincoln Park Conservatory	アメリカ・イリノイ州
Greenwich	イギリス・ロンドン
Greenwich Park	イギリス・ロンドン

表 4 音楽番号 59978 が選曲された観光地と観光地の種類.

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	協会/組織
Broadway	街並み
Radio City Music Hall	劇場
Le Puy du Fou	テーマパーク
Westminster	街並み

地レビューをベクトル化することで、観光地レビューを歌詞として解釈した。そして、ベクトル間の類似度を参照することで、観光地レビューと類似した歌詞を選曲した。比較手法として用意したレビューコーパスを学習した単語分散表現モデルの選曲結果について考察したところ、どちらのモデルを用いた場合でも観光地の性質が推薦結果に影響を与えることが確認された。

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。また、選曲された歌詞についての定性的な考察も

行っていく。選曲システムとしては、観光地に対する選曲結果に対して更にインタラクティブに選曲を促すような探索的な検索アルゴリズムについても考えていく。

謝辞

本研究は、一部、科研費若手 B#16K21482, および、挑戦的萌芽研究#15K12151 の支援のもと行われた。本研究の歌詞データの収集には、豊橋技術科学大学吉田光男助教の協力を得た。記して、謝意を表す。

参考文献

- [1] Steffen Pauws. Cubyhum: A fully operational query by humming system. In *Proceedings of ISMIR 2002*, pages 187–196, 2002.
- [2] Mitsunori Ogihara, See Profile, Dingding Wang, Tao Li, and Mitsunori Ogihara. Are tags better than audio features? the effect of joint use of tags and audio content features for artistic style clustering, 2016.
- [3] Elias Pampalk, Arthur Flexer, and Gerhard Widmer. Improvements of audio-based music similarity and genre classification. In *Proceedings of the 6th International Conference on Music Information Retrieval*, 2005.
- [4] Renato Panda, Ricardo Malheiro, and Rui Pedro Paiva. Musical texture and expressivity features for music emotion recognition. In *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018, Paris, France*,

- September 23-27, 2018, pages 383–391, 2018.
- [5] Kosetsu Tsukuda, Keisuke Ishida, and Masataka Goto. A lyrics-based music exploratory web service by modeling lyrics generative process. In *Proceedings of the 18th International Conference on Music Information Retrieval*, page 544 - 551, 2017.
 - [6] Rui Cai, Chao Zhang, Shong Wang, Lei Zhang, and Wei-Ying Ma. Musicsense: contextual music recommendation using emotional allocation modeling. In *Proceedings of the 15th ACM international conference on Multimedia*, page 553 - 556, 2007.
 - [7] Marius Kaminskas, Francesco Ricci, and Markus Schedl. Locationaware music recommendation using auto-tagging and hybrid matching. In *Proceedings of the 7th ACM conference on Recommender systems*, page 17 - 24, 2013.
 - [8] Matthias Braunhofer and Marius Kaminskas and Francesco Ricci. Location-aware music recommendation. *International Journal of Multimedia Information Retrieval*, 2(1):31–44, 2013.
 - [9] Ignacio Fernández-Tobias, Iván Cantador, Marius Kaminskas, and Francesco Ricci. Cross-domain recommender systems: A survey of the state of the art. In *Proceedings of the 2nd Spanish Conference on Information Retrieval*, page 187 - 198, 2012.
 - [10] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 851–860, 2010.
 - [11] Eiji Aramaki, Sachiko Maskawa, and Mizuki Morita. Twitter catches the flu: Detecting influenza epidemics using twitter. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1568–1576, 2011.
 - [12] Sheeba Naz, Aditi Sharan, and Nidhi Malik. Sentiment classification on twitter data using support vector machine. In *Proceedings of 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 676–679, 2018.
 - [13] Mitsuo Yoshida and Fujio Toriumi. Information diffusion power of political party twitter accounts during japan’s 2017 election. In *Proceedings of Social Informatics - 10th International Conference, SocInfo 2018, St. Petersburg, Russia*, pages 334–342, 2018.
 - [14] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Proceedings of HLT-NAACL*, pages 746–751, 2013.
 - [15] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.
 - [16] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc., 2013.
 - [17] Marius Kaminskas and Francesco Ricci. Emotion-based matching of music to places. In Jan Fagerberg, David C. Mowery, and Richard R. Nelson, editors, *Emotions and Personality in Personalized Services*, pages 287–310. Springer, 2016.