

## 状況に応じた楽曲推薦に向けたソーシャルデータ分析

## Social Data Analysis for Situation-aware Music Recommendation

北谷 光希 高間 康史  
Mitsuki Kitaya Yasufumi Takama

首都大学東京大学院システムデザイン研究科  
Graduate School of System Design, Tokyo Metropolitan University

This paper aims to extract relationship between songs and their playback situation from social data. Target data are tweets containing “#nowplaying” hashtag, from which the song title and singer name as well as the situation at the time of playing songs such as “rain”, “night” are extracted. Based on the extracted information, we build a music database that stores songs and corresponding situations. Through the construction of the music database, this paper discusses the situation at the time of playing songs that can be acquired from tweets.

## 1. はじめに

本稿では、楽曲とその再生時の状況の関係をソーシャルデータから分析する手法を提案する。

現在、ポータブル mp3 プレイヤや音楽配信サービスの普及により数多くの楽曲を容易に楽しむことが可能となった。その結果、好みの楽曲を手早く再生したり未知の楽曲と出会うことができる楽曲推薦システムの需要が高まり、重要な研究テーマとなっている [竹川 08][古屋 15][津谷 13][梶 04][新美 11]。

従来の楽曲推薦システムでは、ユーザの再生履歴に基づく協調フィルタリングや、音響的特徴や歌詞といった内容ベースの推薦が行われている。これらに対し提案手法ではソーシャルデータに着目し、状況に応じた楽曲推薦を目的としてソーシャルデータの分析を行う。SNS の Twitter\*1 に投稿されたハッシュタグ「#nowplaying」付きツイートを収集することにより、アーティスト名と楽曲名および楽曲再生時の状況を抽出する。抽出した情報に基づき、楽曲とそれに適した状況の関係を楽曲データベースとして構築する。楽曲データベースの構築を通じて、ツイートから取得可能な楽曲再生時の状況について分析する。

## 2. 関連研究

情報推薦システムとはユーザの興味や嗜好に合わせたアイテムをユーザに提供するシステムである。これによりユーザは膨大な情報や商品の中から必要なアイテムを入手することができ、オンラインショッピングサイトや音楽配信サイトなどで広く応用されている。情報推薦の手法は大きく分けて内容ベースフィルタリングと協調フィルタリングに分けられる [神島 07]。

楽曲推薦システムではユーザが好む楽曲または状況や気分に応じた楽曲を推薦する。前者ではユーザの楽曲に対する再生履歴や評価などに基づいて推薦を行い、後者ではユーザの現在の状況や気分をユーザ自身が入力したりハードウェアから自動的に取得したりすることで、それに合わせた推薦を行う。また、楽曲推薦システムにおいては推薦精度だけでなく、意外性のあるアイテムを推薦することが重要であると言われてい

る。そのため、アーティスト名や楽曲名に拘らずに未知の楽曲と出会うことができる「状況に応じた楽曲推薦」は有力な手法となっている。推薦手法としては、テンポや音域などの音響的特徴 [竹川 08]、歌詞情報 [古屋 15]、「明るい」、「静かな」といった楽曲に対する印象語 [津谷 13]、ユーザの楽曲再生時の状況 [梶 04] といった内容ベースフィルタリングを用いたものや、協調フィルタリング [新美 11] を用いたものなどが提案されている。

## 3. 提案手法

図 1 に、本稿における分析手順を示す。Twitter からハッシュタグ「#nowplaying」付きのツイートを収集し、それらからアーティスト名、楽曲名、楽曲再生時の状況を取得する。取得した情報に基づき、楽曲と状況を関連付けて楽曲データベースに格納する。以下に各処理の詳細を述べる。



図 1: ツイートの分析手順

### 3.1 ツイートの収集

ツイートの収集には Twitter Search API\*2 を用いた。Twitter Search API により、「#nowplaying」付きツイート本文および投稿日時を収集する。

「#nowplaying」付きツイートを収集し分析したところ、アーティスト名、楽曲名のみで投稿されることが多く、そのままでは楽曲再生時の状況をほとんど取得することができないことが判明した。そこで提案手法では、「#nowplaying」付きツイートを収集する際に、同一投稿者により前後 5 分以内に投稿されたツイートも収集する。これは投稿者の状況は連続したツイート内に出現するという仮定に基づいたものであり、本稿では `branch_tweet` と呼ぶ。

### 3.2 ツイートからの楽曲情報抽出

収集したツイートからアーティスト名、楽曲名を抽出する。Twitter におけるツイートは一般的に口語的な表現が用いられ

連絡先: 高間康史, 首都大学東京大学院システムデザイン研究科, 〒 191-0065 東京都日野市旭が丘 6-6, ytakama@tmu.ac.jp

\*1 <https://twitter.com/>

\*2 <https://dev.twitter.com/rest/public/search>

るため、そのままでは必要情報を抽出することは困難である。また、楽曲名のみでアーティスト名が書かれていない場合なども多い。よってツイートから抽出したテキストに基づき楽曲情報取得 API を用いて検索を行い、情報を補完する。そのため、前処理として収集したツイートを楽曲検索に適した形式に整形する。

### 3.3 楽曲再生状況の抽出

整形されたツイートから、抽出したアーティスト名、楽曲名を除去した残りの部分を本研究では remnant と呼び、これから楽曲再生時の状況を抽出する。remnant に branch\_tweet を付帯させたものを remnant' として処理対象とする。remnant' を整形したのち、状況に関する単語を格納した situation\_dict と比較することによって楽曲再生時の状況を取得する。situation\_dict は、他の楽曲推薦システム [津谷 13][梶 04]<sup>\*3\*4</sup> を参考に独自に作成した辞書であり、106 種類の状況に関して合計 378 語を格納している。

また、時間に関する状況として、ツイート投稿日時から時間、曜日、月も取得する。

### 3.4 データベースの作成

取得できたすべての楽曲再生時の状況を楽曲と関連づけてデータベースを作成する。データベースの形式として RDF[W3C Working Group] による記述を採用した。主語となるリソースはアーティスト名と楽曲名を連結させ、URL エンコードしたものとした。このリソースを一つの楽曲とし、アーティスト名、楽曲名、状況をそれぞれ関係付ける。

## 4. 分析結果

表 1: 収集ツイートの分析結果概要 (件)

楽曲検索を行ったツイート	66879
楽曲情報が取得できたツイート	57904
remnant' が存在したツイート	33883
remnant' が存在したツイート (branch_tweet 不使用)	24877
状況が取得できたツイート	4975
状況が取得できたツイート (branch_tweet 不使用)	1591

2016 年 8 月 10 日から 2016 年 9 月 3 日の間に投稿されたツイートを収集し、合計 66879 件のツイートについて分析・評価を行った。結果を表 1 にまとめる。楽曲情報取得 API での検索によってアーティスト名、楽曲名を取得できたツイートが 57904 件あったが、この中には誤判定されたものも含まれる。9 月 1 日に投稿された 580 件のツイートに対し人手で誤判定を集計した結果、誤りは 41 件であった。この結果から、母比率の推定により信頼度 95% で正解率を求めると 91%~95% となり、提案手法により効率的に楽曲情報が収集可能であることがわかる。誤判定されたツイートを確認すると、使用した API のデータベース内の誤情報、ツイートの整形ミス、ツイート取得時のフィルタリング機能の不足などが原因であった。

remnant' が存在したツイートのうち、状況が取得できたツイートは 4975 件 (7%) であり、楽曲情報と比較して取得が困難であることがわかる。状況が取得できなかったツイートを確認すると、楽曲再生時の状況に関する文章が付属していない場合があった。また、branch\_tweet が存在した場合でも、楽曲とは関係のない内容が書かれている場合が多く存在した。

branch\_tweet の有無による状況の取得数の差を比較すると、branch\_tweet を使用した場合に状況を取得できたツイート数

は 4975 件であったのに対し、使用しなかった場合には 1591 件であった。この結果から branch\_tweet が楽曲と状況を関連付ける上でデータ量の増加に貢献していることがわかる。

表 2 に状況が取得できたツイートの例を示す。「雨」、「作業」など、楽曲再生時の状況が正しく取得できていることから、提案手法が有効であることがわかる。

表 2: 状況が取得できたツイートの例

ツイート	もうすっかり雨もあがったけど Adventure - [Alexandros] #NowPlaying
branch_tweet	思ったより全然早く仕事終わっちゃった、Swan 店着朝 7 時なのになー(笑) 明日実家帰る前に買っ てくか帰ってくる時に買うか、悩む(´ω´)
artist	[Alexandros]
title	Adventure
取得した状況	雨, 作業, 朝, 帰り

## 5. おわりに

本稿では、楽曲とその再生時の状況の関係をソーシャルデータから分析する手法を提案した。収集したツイートを分析した結果、高い割合で楽曲情報の抽出が可能であることを示した。また、ツイートから楽曲と再生時の状況との関係を取得可能であることを示したが、取得できたツイートの割合は低く、分析手法に改善の余地があることも判明した。今後は取得情報量増加のためのシステムの改善や、作成したデータベースを用いた楽曲推薦システムの開発などを旨とする。

## 参考文献

- [竹川 08] 竹川和毅, 土方嘉徳, 西田正吾, 内容に基づく音楽データの探索・推薦システム, 人工知能学会論文誌, Vol.23, No.5, pp.330-343, 2008
- [古屋 15] 古屋瑞生, 黄宏軒, 川越恭二, 聴取目的に応じた音楽推薦のための歌詞からの音楽印象分類方法, DEIM Forum 2015, B8-1, 2015
- [津谷 13] 津谷篤, 西牧駿矢, 小見裕明, 田中敦, 印象の有無のみの感性評定を用いた状況に応じた楽曲選択法とその応用, 日本感性工学会論文誌, Vol.12, No.3, pp.397-407, 2013
- [梶 04] 梶克彦, 平田圭二, 長尾確, 状況と嗜好に関するアノテーションに基づくオンライン楽曲推薦システム, 情報処理学会 音楽情報科学 (MUS) 研究会, Vol.2004, No.127(2004-MUS-058), pp.33-38, 2004
- [新美 11] 新美怜志, 濱川礼, 協調フィルタリングを用いた音楽推薦とマッピング手法, 情報処理学会 音楽情報科学 (MUS) 研究会, Vol.2011, No.21(2011-MUS-89), pp.1-6, 2011
- [神寫 07] 神寫敏弘, 推薦システムのアルゴリズム (1), 人工知能学会誌, Vol.22, No.6, pp.826-837, 2007
- [W3C Working Group] W3C Working Group, RDF 1.1 Primer, <https://www.w3.org/TR/rdf11-primer/>, (2017/01/19 現在)

\*3 <https://awa.fm/>

\*4 <http://www.apple.com/jp/music/>