

確率的潜在意味解析と最大事後確率推定を用いた日本語歌詞の印象推定手法の評価

ー 感性空間を用いたユーザの気分合う楽曲推薦システムの開発に向けて ー

学籍番号 18118047
氏名 黒川 皇輝
指導教員 宮治 裕 教授

1. はじめに

近年 Amazon Music, Spotify といった音楽配信サービスとスマートフォンなどのモバイル端末の普及により、人々は膨大な楽曲を手軽に聴くことができるようになった。それに伴い音楽配信サービスのユーザは増加傾向にある [1]。膨大な楽曲の中からユーザが自身の嗜好に合う楽曲を見つけるのは困難であるため、楽曲を推薦するシステムが研究トピックとして着目されている。

ユーザの嗜好は時間と共に移り変わる。したがって、ユーザの気分や状況に合う楽曲を推薦するコンテキストウェア楽曲推薦システムの需要が存在する。コンテキストは次のように定義されている。「エンティティの状況特徴化するのに用いられるあらゆる情報。エンティティとは、ユーザとアプリケーションとのインタラクションに関連する人や場所、オブジェクトを指し、それにはユーザ自身とアプリケーション自体も含まれる」[2]。

テキストや映像、画像などの音楽以外のメディアからユーザが得る感性情報をマルチメディアコンテキストという。マルチメディアコンテキストの観点から、歌詞からユーザが受ける印象を推定する研究が行われている。しかし、いまだ確立された手法は存在していない。

本研究では日本語歌詞から読み取れる印象を推定する手法の設計と評価をする。

2. 関連研究と本研究の位置付け

角田ら (2018) [3] は印象を Plutchik の感情モデルの基本 8 感情 [4] を用いて推定した。形態素解析ツールを用いて歌詞から単語を抽出する。その後 Word2Vec を利用してのベクトル変換を行ない、基本 8 感情との類似度を算出して、一番類似度が高い感情をテキストから得られる印象とした。

大木ら (2018) [5] は日本語評価極性辞書にのっとり、歌詞をポジティブ・ネガティブ・ニュートラルの 3 つの印象に分類した。3 つの評価極性を歌詞内の各単語に付与し、単語の評価極性値を平均化することで歌詞自体への印象の評価値を算出した。

歌詞情報から印象を推定する研究は未だ発展途上であるため、Word2Vec や日本語評価極性辞書以外で印象推定する手法を提案する。西川ら [6] は英語の感情価単語セット ANEW [7] を用いて、歌詞を Russell の

Arousal-Valence 平面 [8] (以後 AV 平面) 上で表現する研究をした。歌詞の各フレーズが持つ単語の AV 平面座標を推定して、フレーズごとの平面座標を求めた。この印象推定手法は英語の歌詞を用いて効果測定が行われたが、日本語の歌詞での効果測定は行われていないので、日本語の歌詞の場合でもこの手法が効果を発揮するのか調査する。本間ら [9] は ANEW の単語を参考にして、日本語訳した単語セットを作成した。よって、この日本語の単語セットを利用すれば、西川らが考案した手法を日本語の歌詞の印象推定でも一定の効果を挙げられることが予想される。

3. 歌詞印象推定手法

3.1. 印象の定義

本稿で推定する印象とは Russell の AV 平面上で表現する。AV 平面は縦軸に Arousal を、横軸に Valence を取る 2 次元平面である。Arousal 軸は正の方向に興奮を、負の方向に弛緩を表す。そして、Valence 軸は正の方向に快を、負の方向に不快を表す。AV 平面の 4 象限は第 1 象限から順に喜怒哀楽の印象を表す。

3.2. 歌詞データの収集

歌詞データの収集は Uta-Net^{*1} から人気のアーティストの曲を 6,813 曲収集した。本研究では 6,813 曲から 3,000 曲を無作為に選んだ。

3.3. ANEW データの拡張

本間らが開発した日本語版 ANEW の単語の類義語と同義語を WordNet [10] を用いて探索する。発見した単語に類義語元の単語が持つ Arousal と Valence の値を与えることで、日本語版 ANEW を 14,232 語に拡張した単語データセットを日本語版 ANEW 拡張データセットとする。日本語版 ANEW 拡張データセットの構成は A+V+ 平面に 5,581 語、A+V- 平面に 2,681 語、A-V- 平面に 3,456 語、A-V+ 平面に 2,493 語である。

3.4. 歌詞データの整形

収集した歌詞データの中に一部英語が入っていたので、英語の歌詞を削除した。そして歌詞データを歌詞のフレーズごとに分解した。形態素解析器 MeCab を用いて各フレーズから名詞・形容詞・動詞を抜き出した。歌

^{*1} <https://www.uta-net.com/>

詞から抜き出した単語の総数は 109,450 語である。

3.5. 歌詞印象推定

確率的潜在意味解析 (PLSA) を利用する。歌詞のフレーズを文書 d とし、フレーズ中に出現する単語を単語 w と定義してモデルパラメータを推定した。 $P(w|z)$ はトピックから単語が観測される確率であるため、この確率の高い単語がトピックを表現する。歌詞の印象を推定するためには潜在的なトピック z を印象に制限する必要がある。通常の PLSA は文書と単語の共起確率に着目して、潜在的なトピックを推定する手法であるため、必ず潜在的なトピックが印象を表現するとは限らない。よって、日本語版 ANEW 拡張データセットを事前知識として使用し、モデルパラメータを MAP 推定することで潜在的なトピックに AV 平面の各象限を表現する。具体的に潜在的なトピック z を次の式 (1) のように AV 平面の各印象として定義する。

$$z \in \{A + V +, A + V -, A - V -, A - V +\} \quad (1)$$

モデルパラメータ $P(w|z)$ の対数事前分布を共役事前分布を用いて式 (2) に定義する。 k は出現する単語の集合を表す。

$$\log P(\theta) \propto \sum_k \sum_z (\alpha_{w_k, z} - 1) \log P(w_k|z) \quad (2)$$

共役事前分布のハイパーパラメータ α は日本語版 ANEW 拡張データセットに含まれる単語で該当の象限に位置する場合のみ原点からの距離を入力する。それ以外の場合無情報事前分布を与える。ラグランジュの未定乗数法を使用して対数尤度関数と対数事前分布より MAP 推定値を求める MAP 推定値は式 (3) で表す。

$$P(w_k|z)_{\text{map}} = \frac{\alpha_{w_k, z} - 1 + p(w_k|z)}{1 - K + \sum_k \sum_z \alpha_{w_k, z}} \quad (3)$$

$P(z)$ と $P(d|z)$ の推定には無情報事前分布を与えた。ベイズの定理より $P(z|w)$ は式 (4) で求められる。

$$P(z | w_k) = \frac{P(w_k | z)_{\text{map}} P(z)}{P(w_k)} \quad (4)$$

各フレーズ d の Arousal と Valence の値は求めた $P(z|w)$ を用いて求めた各単語 w の Arousal と Valence の値を合計、正規化して求める。以上で歌詞中のフレーズごとに感情価を求める。

4. 実験

4.1. 方法

本研究の印象推定方法によって推定された歌詞と歌詞中に出現するフレーズの印象が妥当であるか効果进行评估する。実験参加者は 10 代後半から 20 代前半の大学生男女 11 名である。男女の内訳は男子 7 名、女子 4 名で

ある。実験参加者には実験前に印象の定義について十分に説明をした。

アンケートを作成しデータを収集した。歌詞とフレーズに対して喜怒哀楽いずれの感性を感じるか質問した。質問の解答項目は喜・怒・哀・楽・わからないである。アンケートに使用した歌詞とフレーズは 4 印象のグループからそれぞれ原点からの距離に基づいて ward 法で 3 つのグループにクラスタリングをしたものから 1 つずつランダムで選出した。

4.2. 結果

推定した印象と実験参加者の回答の過半数以上が合っていることを妥当であると定義する。歌詞においては喜のクラスの印象の推定結果が妥当であることを確認した。フレーズにおいては喜のクラスと哀、楽のクラスの 1 部分において妥当であることを確認した。妥当であると判断したグループ数を印象ごとに図 1 で示す。

4.3. 考察

歌詞とフレーズどちらにおいても喜の印象の推定結果が妥当であった理由は事前知識が他の印象と比べて豊富であったためであると考えられる。事前知識が十分に収集できないと MAP 推定する際に未知の単語が多くなる。未知の単語に対して MAP 推定する際には無情報事前分布を与えるので、MAP 推定前の $P(w|z)$ が MAP 推定後の $P(w|z)$ の分布に大きく影響を残す。したがって潜在的トピックを印象に制限できないため妥当でない結果となったと考える。

5. おわりに

本研究では日本語歌詞から読み取れる印象を推定する手法の設計と評価をした。実験の結果喜の印象において推定結果は妥当であることが判明した。しかし、全体的に推定の精度は悪かった。主な理由としては MAP 推定に用いる事前知識が不足していたためであると考えられる。今後の改善点として事前知識を十分に用意するために、利用する単語セットまたは単語セットの拡張方法の改善が必要である。

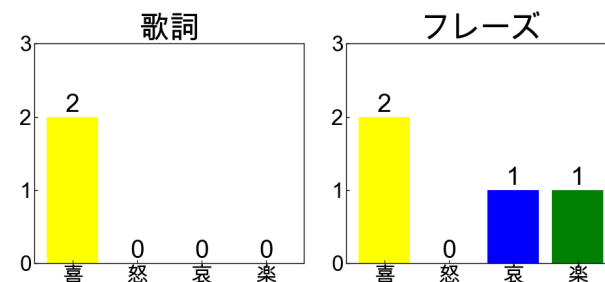


図 1: 妥当であると判断されたグループ数

参考文献

- [1] 一般社団法人日本レコード協会. 2020 年度「音楽メディアユーザー実態調査」.
<https://www.riaj.or.jp/f/pdf/report/mediauser/softuser2020.pdf> (参照 2021-12-17).
- [2] 奥健太. 楽曲推薦システム —プレイリスト, コンテキスト, インタクション—. 人工知能, No. 34, pp. 300–308, may 2019.
- [3] 角田拓己, 大園忠親, 新谷虎松. 対話によるムード推定に基づく楽曲推薦エージェントの開発. 第 80 回全国大会講演論文集, Vol. 2018, No. 1, pp. 157–158, mar 2018.
- [4] Robert Plutchik. A psychoevolutionary theory of emotions. *Social Science Information*, Vol. Vol.21, No. 4-5, pp. 529–553, July 1982.
- [5] 大木麻里衣, 丸野由希, 久保孝富. 歌詞解析と心拍変動分析を用いた楽曲による感情への影響の予備的調査. 2018 年度 情報処理学会関西支部 支部大会講演論文集, No. 2018, sep 2018.
- [6] 西川直毅, 糸山克寿, 藤原弘将, 後藤真孝, 尾形哲也, 奥乃博. 歌詞と音響特徴量を用いた楽曲印象軌跡推定法の設計と評価. 研究報告音楽情報科学 (MUS), Vol. 2011-MUS-91, No. 7, pp. 1–8, jul 2011.
- [7] Margaret M. Bradley, Peter J. Lang, Margaret M. Bradley, and Peter J. Lang. Affective norms for english words (anew): Instruction manual and affective ratings, 1999.
- [8] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, Vol. Vol.39, No. 6, pp. 1161–1178, 1980.
- [9] 本間喜子. 単語の感情価と覚醒度にもとづいた単語刺激の作成. 愛知工業大学研究報告, No. 49, pp. 13–24, mar 2014.
- [10] Francis Bond, Timothy Baldwin, Richard Fothergill, and Kiyotaka Uchimoto. Japanese semcor: A sense-tagged corpus of japanese.