

音楽感情データセットを用いた歌詞に対する感情分析

2020SC110 吉澤莉空

指導教員：河野浩之

1 音楽と歌詞が与える影響

近年、多くの SNS やデバイスにより音楽はより身近になっている。ただ身近になっただけでなく、松本 [1] の研究によって、悲しい音楽の聴取によって悲しみが軽減されることが明らかになっているように、音楽の聴取は人の心理状態に良い影響を与えることが示されている。音楽の聴取によって特定の感情の軽減や誘発が出来る為である。また、それに対する歌詞の影響も研究されており、Brattico ら [2] が脳の活動より、ポジティブな音楽聴取時の感情に対する歌詞の影響は無かったが、ネガティブな音楽では歌詞の影響が見られたとしているように、音楽聴取時に認識される感情に対する歌詞の影響は音響的な部分よりも弱い、ネガティブな感情を認識する音楽に対してより大きな効果をもたらすことが示されている。これらを利用するため、感情に合わせた楽曲提案システムの研究、開発が進められており、音楽全体だけでなく、歌詞単体に対する感情分析とそれを利用した該当システムの研究、開発が行われている。

しかし、歌詞に対する感情分析には言語の壁が存在するため、歌詞感情データセットが無い言語に対しては辞書型の感情分析が行われており、近年大きく発展しつつある大規模言語モデルを活かした、文脈や細かい表現を基にした感情分析が出来ない状況にある。だが、音楽の音響的な部分と歌詞が含まれている楽曲である歌に対する感情アノテーションが行われている音楽感情データセットなら歌詞の影響を含んでいるはずである。よって、本研究では音楽感情データセットを用いて歌詞感情分析モデルを訓練し、歌詞感情データセットで精度評価をすることで音楽感情データセットが歌詞感情分析に利用可能かどうかを調べる。

2 音楽と歌詞の感情分析の関連研究

感情モデルの 1 つである Russell ら [3] の感情円環モデルを用いた、音楽と歌詞に対する感情分析とデータセットについて、本研究で用いるものを紹介する。ラッセルの感情円環モデルとは高い値であるほど興奮しており、低い値であるほど落ち着いているとした arousal 値を縦軸、高い値であるほどポジティブで低い値であるほどネガティブであるとした valence 値を横軸とした 2 次元感情モデルで、28 個の感情点が 2 次元図上に円環を描くことからラッセルの感情円環モデルと呼ばれる。以降、呼称をラッセルの感情モデルとする。

Çano ら [4] は、歌詞に対して感情辞書をベースとした感情価のアノテーションをラッセルの感情モデルに基づいて行い、感情カテゴリのいずれかに分類した MoodyLyrics

データセットを公開した。3 つの単語辞書を合わせた混合辞書を用いてアノテーションを行い、valence 値、arousal 値を決定した。その後しきい値の設定によって、happy, sad, relaxed, angry の 4 つの感情に分類、また当てはまらなかった曲を削除した。既に人間によって歌詞に対する感情のアノテーションが行われていた歌詞感情データセットと比較した結果、データセット全体としての精度は 74.25% であった。

Zhang ら [5] は短く抜粋した曲を聴いた人による EDA(皮膚電気活動) 反応と人によるアノテーションを行い、ラッセルの感情モデルを基にその値を定めた PMEmo データセットを公開した。合計 457 人の被験者に感情のアノテーションを行わせた。動的な感情、それを平均した静的な感情、EDA データ、それぞれについて別で訓練を行った結果、静的な感情における感情分析は高い評価を得た。

Agrawal ら [6] は歌詞に対する感情分析モデルとして、XLNet Transformer モデルと、その出力を Sequence summary によって要約、計算を行うモデルを提案した。MoodyLyrics の 4 つの感情分類だけでなく、arousal 値と valence 値の予測も行うマルチタスクなモデルとしても利用することが出来る。MoodyLyrics で訓練したモデルをテストデータで評価した結果は、4 つの感情分類では 94.78%、arousal 値予測は 95.73%、valence 値予測は 94.38% と高精度であった。

3 歌詞感情分析に適応する提案手法

音楽感情データセットである PMEmo データセットの valence 値と arousal 値の閾値を動かし、MoodyLyrics と同じ 4 つの感情カテゴリに分ける。それを学習した Agrawal ら [6] の感情分類モデルを MoodyLyrics データセットで評価する。この操作により、評価が高くなる閾値を求める。また、Google が提供している Google Colaboratory 上で Python を用いて本研究用のプログラムを作成している。表 1 はデータの概要である。

表 1 分析に用いるデータの概要

データセット	happy	sad	relaxed	angry
MoodyLyrics	657 曲	483 曲	495 曲	507 曲
PMEmo				
valence = 0.5				
arousal = 0.5	386 曲	108 曲	49 曲	59 曲

最終的に、arousal 値 valence 値共に 0.5 の閾値 (初期閾値) を設定した PMEmo データセットを訓練したモデルと、MoodyLyrics に適応させた閾値を設定した PMEmo データセットを訓練したモデルを MoodyLyrics で評価し

たときの精度評価比較を行う。

また、歌詞感情分析に用いるモデルとして先行研究で紹介した Agrawal ら [6] のモデルを使用する。MoodyLyrics を用いた訓練と評価において高い精度を発揮していたことや、簡易的に利用することが出来たこと、また、尺度が同じであるため、評価に MoodyLyrics を用いる本研究においても、このモデルを用いることとした。Agrawal らは token 数を 1024、バッチサイズを 8、テストサイズ 0.2 として訓練をしていたが、計算資源の不足により、本研究ではバッチサイズを 4 として訓練を行っている。また、バッチサイズ 4 で MoodyLyrics を用いて訓練と評価を行った結果、F1 スコア 90.77% と高精度であった。よって、バッチサイズ縮小によるモデルの大きな劣化は無く本研究においても利用可能であると言える。

4 研究結果

総当たりで valence 値 arousal 値それぞれ 0.1 毎に閾値を設定し F1 スコアを出力、比較した結果、valence=0.60, arousal=0.70 のとき、F1 スコアが一番高かった。次に 0.05 ずつ閾値を変更したが、先に提示した閾値の精度評価が一番高かった。よって、valence=0.60, arousal=0.70 を適応閾値 1 とする。また、適応閾値 1 を中心に、再度 0.01 毎に閾値を変更、設定した結果、valence=0.59, arousal=0.68 のとき、F1 スコアが一番高かった。よって、これを適応閾値 2 とした。適応閾値 1, 2 と valence=0.50, arousal=0.50 の初期閾値によって設定、訓練されたモデルの精度評価を比較する。表 2 にて 2 つの適応閾値と初期閾値で訓練したモデルの精度評価を比較する。

表 2 適応閾値と初期閾値の評価比較

	Precision[%]	Recall[%]	F1 スコア
適応閾値 1	28.57	35.32	0.27
適応閾値 2	31.57	34.58	0.28
初期閾値	40.14	36.33	0.26

Accuracy はそれぞれ、適応閾値 1 は 36.27, 適応閾値 2 は 37.86, 初期閾値は 40.24 であった。感情毎にも精度評価を行った。そのうち F1 スコアのみを表 3 に示す。

表 3 適応閾値と初期閾値の F1 スコア比較

	angry	happy	relaxed	sad
適応閾値 1	0.00	0.45	0.18	0.46
適応閾値 2	0.00	0.49	0.16	0.45
初期閾値	0.01	0.53	0.00	0.50

5 本研究のまとめ

PMEmo データセットを用いて歌詞感情分析モデルを訓練し、MoodyLyrics による評価を行うことで PMEmo データセットをはじめとした音楽感情データセットで歌詞

感情分析が可能かどうかを研究した。全体を通して歌詞感情分析に使える精度ではなかった。また、閾値を変更しても精度は低下するだけであった。しかし、閾値を変更してデータ数が増減しても happy と sad の F1 スコアは 0.50 前後と比較的高い評価を維持していた。

今後の課題としては、歌詞感情データセットの作成がデータセットが無い言語に対する歌詞感情分析を可能にすることを前提とする。その上で本研究で行った歌詞に対する多クラス分類だけでなく、本研究で高い精度評価を得ていた happy と sad に対する 2 値分類や、その 2 つの感情に則した arousal と valence の感情強度分析の他、SNS や映画レビュー等、別の感情データセットを用いた歌詞感情分析の精度検証をデータセット毎に行うなど、様々なアプローチからデータセットが無い言語に対する歌詞感情分析を、高い精度で行えるようにする必要があると考える。

参考文献

- [1] 松本じゅん子, “音楽の気分誘導効果に関する実証的研究: 人はなぜ悲しい音楽を聴くのか,” 教育心理学研究, vol.50, no.1, pp.23-32, 2002. DOI:10.5926/jjep1953.50.1.23
- [2] E. Brattico, V. Alluri, B. Bogert, T. Jacobsen, N. Vartiainen, S. Nieminen and M. Tervaniemi, “A functional MRI study of happy and sad emotions in music with and without lyrics,” Front. Psychol., vol.2, no.308, pp.1-16, Dec 2011. DOI:10.3389/fpsyg.2011.00308
- [3] J. A. Russell, “A Circumplex Model of Affect,” Journal of Personality and Social Psychology, vol.39, no.6, pp.1161-1178, Dec 1980. DOI:10.1037/h0077714
- [4] E. Çano and M. Morisio, “MoodyLyrics: A Sentiment Annotated Lyrics Dataset,” ISMSI '17: Proceedings of the 2017 International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence, pp.118-124, Mar 2017. DOI:10.1145/3059336.3059340
- [5] K. Zhang, H. Zhang, S. Li, C. Yang and L. Sun, “The PMEmo Dataset for Music Emotion Recognition,” ICMR '18: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, pp.135-142, Jun 2018. DOI:10.1145/3206025.3206037
- [6] Y. Agrawal, R. G. R. Shanker and V. Alluri, “Transformer-Based Approach Towards Music Emotion Recognition from Lyrics,” Advances in Information Retrieval: 43rd European Conference on IR Research, ECIR 2021, pp.167-175, Mar 2021. DOI:10.1007/978-3-030-72240-1_12