

スマートフォンのセンサデータに基づく ユーザコンテキスト認識型音楽推薦システム

山中翔太^{†1}
慶應義塾大学

浜中智己^{†2}
慶應義塾大学

佐々木航^{†3}
奈良先端科学技術大学院大学

大越匡^{†4}
慶應義塾大学

中澤仁^{†5}
慶應義塾大学

概要

現代のオンライン音楽視聴サービスは膨大な曲へのアクセスを提供しているが、逐次的なユーザの感情状態や行動に基づいた推薦は限定的である。本研究では、スマートフォンセンサデータから推定されるユーザの感情や行動に適応した音楽推薦システムを構築する。ユーザの行動情報は iOS が提供する Activity Recognition API を利用することで収集し、感情情報はスマートフォンセンサデータと主観的感情ラベルから学習されたモデルを構築・推定する。本論文では、システム構築に向けたデータ収集実験の実施と収集されたデータの分析、及び感情状態推定モデルの構築・評価について述べる。データは計 32 名の被験者より収集し、感情推定モデルの精度評価の結果は個人ごとに構築したモデル性能で平均して 86.3% の精度を達成した。

1. 背景

近年、AppleMusic[1] や Spotify[2] などの音楽配信サービスは一般的なものとなっている。その多くは、ユーザの視聴履歴から嗜好に合った音楽を推薦するレコメンド機能を有しているが、主にユーザの長期的な嗜好に焦点を当てているため、短期的な変動や即時の状況に基づく推薦は行われていない。既存手法としてよく用いられる再生履歴に基づく推薦手法や協調フィルタリング手法は、いずれも長期的なユーザの嗜好に応えるものである。逐次的に推薦するシステムの研究も存在するが、その多くは位置情報や時間、天気、SNS の投稿文といった外部環境に依存しており、ユーザの感情状態や日々の行動といった内部コンテキストを活用していないのが一般的である [3][4][5]。感情状態を把

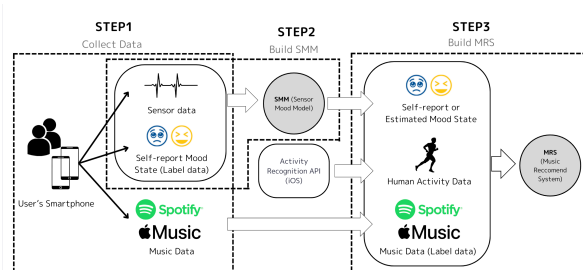


図1 ユーザコンテキスト認識型音楽推薦システムを構成する3つのフェーズ

Figure1 Three Phases that Make up a User Context-aware Music Recommendation System

握するために、特定の生体情報（例えば皮膚電気活動）を取得する研究も進んでいるが、専用の高額な装置が必要とされ、実用性に課題がある。そこで本研究では、ユーザの感情状態と行動をスマートフォンで推定し、逐次的に音楽が推薦されるシステムを提案する。本システムは、既存手法の内面状態把握特性、実用性、汎用性の課題を解決する。実装は大きく分けて3つのステップに分けられる。ステップ1ではスマートフォンに内蔵されたセンサデータの収集と自己申告による主観的感情データの収集、及びユーザの音楽視聴ログを Last.fm API[6] を利用して収集する。ステップ2では選択されたセンサデータの特徴量を使用して機械学習モデルを訓練し、ユーザの感情状態（ポジティブまたはネガティブ）を推定する [7]。ステップ3では、ユーザの行動データ、推定感情状態データ、及び音楽視聴ログを組み合わせることでレコメンドモデルを構築する。本研究のモデル構築フローを図1に示す。

本論文では、ステップ1とステップ2について述べる。

2. 関連研究

Wei-Po Lee らの研究 [8] では、行動認識のみに基づく逐次的な音楽推薦システムを提案している。本研究の課題は同じ行動データの場合においても、異なる感情状態を認識

Smartphone-based activity and emotion-aware music streaming recommendation system

^{†1} SHOTA YAMANAKA, Keio University

^{†2} SATOKI HAMANAKA, Keio University

^{†3} WATARU SASAKI, Nara Institute of Science and Technology

^{†4} TADASHI OKOSHI, Keio University

^{†5} JIN NAKAZAWA, Keio University

できないことである。例えば、個人が家に滞在している具体的な理由 (気分の落ち込みによるものか、あるいは単なる休養のためか) を識別できない。同様に、Jae Sik Lee らの研究 [5] では、季節、月、曜日、天気、温度などの外部環境のコンテキストデータを基に推薦を行なっているが、ユーザ内部のコンテキストデータを考慮していないという課題がある。Web テキストを読む状況に絞った音楽推薦の手法 [9] や、観光スポットに基づいた手法 [4] など提案されているが、いずれもユーザが特定の状況下にある場合のみ機能するため、汎用性に課題が残る。本研究では、スマートフォンのセンサデータを利用してユーザの日常的な感情や行動を推定し、逐次的に音楽を推薦するシステムを構築する。

3. データ収集実験

感情推定モデル、及びユーザコンテキストに基づく音楽レコメンドモデルの構築のためのデータ収集実験を実施した。実験期間は 39 日間、被験者は 40 人である。被験者は 20 代が 8 割その他 2 割となっている。3.3 章で後述する API の特性により Spotify、もしくは AppleMusic を利用しているユーザに限定した。収集に必要な初期設定はスライド形式のマニュアルによって被験者各自で行ってもらった。収集するデータはスマートフォンセンサデータと主観的感情データと音楽視聴ログの 3 種類である。

3.1. センサデータの収集

収集したセンサデータは 11 種類 (accelerometer, activity, barometer, battery, Bluetooth, gravity, gyroscope, location, rotation, screen, weather) である。これらのセンサデータは、センサデータ収集プラットフォームである AWARE Client V2[10, 11](以下 AWARE) を用いて収集した。AWARE によってクライアントデバイスのバックグラウンドで収集されたセンサデータは、デバイスの WiFi 接続時、及び電源接続時にデータベースサーバに送られ、保存される。センサデータのサーバへの未送信を防止するため、被験者マニュアルに AWARE アプリケーションのタスク終了を避けるよう指示を記載した。これにより、データの未送信を防ぐ。

3.2. 自己報告による主観的感情データの収集

感情推定モデルを構築する上で必要な正解ラベルデータとなる主観的感情データを収集した。感情データは ESM (Experience Sampling Method) 手法 [12] を用いたアンケートを 1 日に 6 回、3 時間おきにユーザのスマートフォンに通知する。ESM による主観的感情ラベルを収集する手法として PAM(Photograph Affect Meter)[13] を採用し、直感的にユーザの気分に合う画像を選択することで主観的感情情報を収集した。ESM のスクリーンショットを図 2 に示



図 2 センサ収集画面 (左) と感情値収集画面 (右)

Figure 2 Sensor collection screen(left) and emotion value collection screen(right)

す。アンケートは回答数確保のため定期的にユーザにリマインドをした。

3.3. 音楽視聴ログの収集

レコメンドモデル作成にあたり音楽視聴ログも収集した。ログは Last.fm[6] という音楽解析アプリの API を用いて、タイムスタンプ、曲名、アーティスト名の情報を収集した。各楽曲のメタ情報を収集するため、Spotify の API によって各楽曲に割り当てられた複数の音楽メタ情報 (length, popularity, danceability, acousticness, energy, instrumentality, mode, liveness, loudness, speechiness, tempo, timesignature) を収集した。

2 つの異なる API を用いた理由は API の特性にある。レコメンドモデルを作成する際、ユーザのコンテキストデータと音楽視聴ログを同じ時間軸で同期させる必要があるため、音楽のタイムスタンプの取得が必須となる。Spotify の API は、他の音楽配信サービスと比較して多様なパラメータを提供しているが、タイムスタンプの情報は提供していない。そのため Last.fm の API を利用し、タイムスタンプ、取得した後に Spotify の API から楽曲の数値情報を検索するという手法をとった。本データ収集実験では、Last.fm の API は Spotify と AppleMusic のみ対応しているため被験者は 2 つのサービスユーザに制限した。

Negative		Positive	
High Negative Affect		High Positive Affect	
Afraid (-8)	Tense (-7)	Excited (+8)	Delighted (+7)
Frustrated (-6)	Angry (-5)	Happy (+6)	Glad (+5)
Low Negative Affect		Low Positive Affect	
Miserable (-4)	Sad (-3)	Calm (+4)	Satisfied (+3)
Gloomy (-2)	Tired (-1)	Sleepy (+2)	Serene (+1)

図 3 ラベル定義の仕方

Figure 3 Label Definition Method

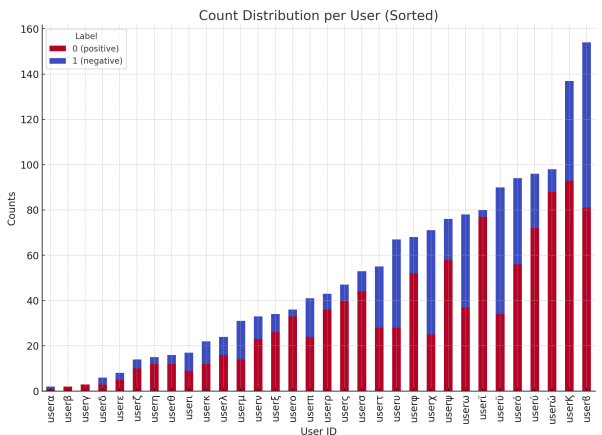


図 4 ユーザごとの感情値の収集数と感情ラベルの割合
Figure4 Number of Emotional Values Collected per User and Percentage of Negative or Positive

4. 収集したデータの分析

4.1. 感情データの分析

被験者 40 名の感情値のアンケートは、最大回答数 9360 のところ 1741 回答、回答率は 18.6% となった。また、一回以上アンケートに回答した人数は 32 名であった。被験者の 2 値感情値 (ネガティブ or ポジティブ) の回答割合はネガティブ 34.57%, ポジティブ 65.43% となった。図 4 にユーザごとの 2 値感情値の回答割合と量を示す。2 値分類は PAM[13] によって取得された感情値 16 種を、valence 値の正負でラベル定義を行った。図 3 は、感情値の分類表である。横軸が valence 値 (Negative or Positive), 縦軸が arousal 値 (High or Low) である。本モデルでは横軸の正負でラベル定義を行った。

4.2. 音楽視聴ログの分析

音楽視聴ログの取得ができたのは 40 名中 25 名であった。ユーザごとに取得できた楽曲数の平均曲数は 554.72 曲, 標準偏差は 512.71 曲, 最小値は 10 曲, 最大値は 2381 曲となった。全ユーザの合計は 16851 曲となった。図 5 にユーザごとの収集曲数と楽曲の valence 値の割合を示す。valence 値

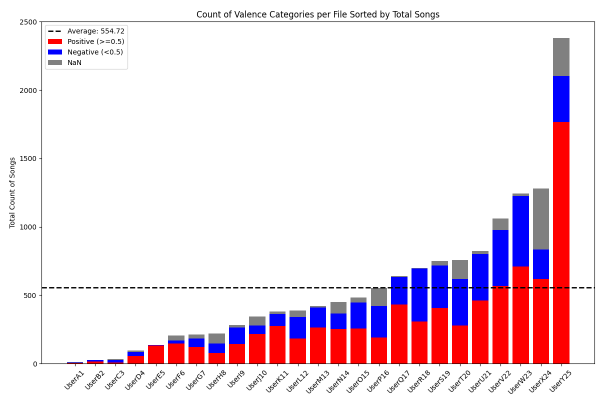


図 5 ユーザごとの曲数とラベル定義した valence 値の割合
Figure5 Number of Songs per User and Percentage of label-defined valence values

は Spotify[2] の API が提供する値であり, 0.0 1.0 の範囲である。図 5 では 0.5 未満を negative, 0.5 以上を positive として色分けして表現した。

5. 感情推定モデルの構築と評価

本研究では、スマートフォンから収集したセンサデータを用いて、ユーザの感情状態 (ポジティブ/ネガティブ) を推定するモデルを構築した。モデルのアーキテクチャとして XGBoost を採用し、2 値分類問題として扱った。

データセットの構築にあたって、ESM による感情状態の回答時刻を中心に、前後 1 時間のセンサデータを抽出した。この 2 時間の時間枠に対して、10 分間の窓を 5 分ずつスライドさせながら特徴量を算出することで、一つの ESM 回答に対して複数の特徴量セットを生成した。これにより、感情状態の推定に適した時間的な粒度を確保しつつ、十分な学習データ数を確保することが可能となった。感情推定の基盤となる特徴量一つの 10 分の時間窓に対して、加速度計・ジャイロスコプから得られる 3 軸データ (x, y, z) と気圧計データを用いて多角的に抽出した。

モデルの構築にあたっては、個人差を考慮するためにデバイス ID ごとに独立したパーソナルモデルを訓練する方式を採用した。図 6 で示す通り、個人ごとに構築したモデル性能は平均して 86.3% の精度を達成し、中央値は 86.1% であった。デバイス間での精度にはある程度のばらつきが見られ、最も低いケースで約 75%, 最も高いケースで約 97% の精度を示した。これは個人の行動パターンや感情表現の違いを反映していると考えられる。

6. 考察

本研究の課題は、センサデータ収集実験の際に被験者がバックグラウンドで起動している AWARE をタスクキルす

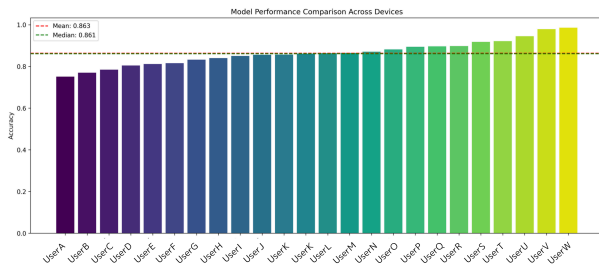


図6 ユーザごとの性能評価結果

Figure6 Performance Evaluation Results by User

ることで、ユーザ単位でセンサデータの欠損値が多いことである。3.1 章で取得したセンサデータの多くが、アンケート回答のためスマートフォンを取り出すタイミングに限られている。現状の感情推定モデルは以上の特定のコンテキストに絞った学習データによる推定のため、ユーザの日常的な生活を反映していない。今後は学習に必要なデータペア数に達していないユーザに対して、再実験を実施し、十分なセンサデータの量を確保する予定である。

7. 結論と今後の展望

本論文では、ステップ1のデータ収集実験、及びステップ2の感情推定モデルの作成まで実施した。ステップ1ではスマートフォンに内蔵されたセンサデータの収集、自己申告による主観的感情データの収集、及びユーザの音楽の視聴ログを Last.fm API[6] を利用し収集した。ステップ2では選択されたセンサデータの特徴量を使用して機械学習モデルを訓練し、ユーザの感情状態を推定 [7] した。今後はステップ3として、ユーザの行動データ、推定感情状態データと音楽視聴ログを組み合わせてレコメンドモデルを構築する予定である。レコメンドモデルは、感情推定モデルより出力された感情値、及び iOS の Activity Recognition API による行動認識データを入力に推薦音楽情報が出力されるモデルである。データ収集実験にて収集した音楽視聴ログの値をラベルデータに、機械学習モデルを作成する。利用する音楽の値は Spotify の API の提供する valence 値を採用する予定である。

謝辞

本研究は、JST CREST JPMJCR19A4, JST RISTEX JPMJRX21J1 の支援を受けたものである。本研究は、JSPS 科研費 JP24K02935 の助成を受けたものである。

参考文献

- [1] Inc., A.: Apple Music, <https://www.apple.com/jp/apple-music/>. Accessed: 2024-12-06.
- [2] AB, S.: Spotify, <https://www.spotify.com>. Accessed: 2024-12-06.

- [3] Kaminskas, M. and Ricci, F.: Location-adapted music recommendation using tags, *User Modeling, Adaptation and Personalization: 19th International Conference, UMAP 2011, Girona, Spain, July 11-15, 2011. Proceedings 19*, Springer, pp. 183–194 (2011).
- [4] Ankolekar, A. and Sandholm, T.: Foxtrot: a soundtrack for where you are, *Proceedings of Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications*, pp. 26–31 (2011).
- [5] Lee, J. S. and Lee, J. C.: Music for my mood: A music recommendation system based on context reasoning, *Smart Sensing and Context: First European Conference, EuroSSC 2006 Enschede, Netherlands, October 25-27, 2006 Proceedings 1*, Springer, pp. 190–203 (2006).
- [6] Ltd., L.: Last.fm, <https://www.last.fm>. Accessed: 2024-12-06.
- [7] Sasaki, W., Hamanaka, S., Miyahara, S., Tsubouchi, K., Nakazawa, J. and Okoshi, T.: Large-Scale Estimation and Analysis of Web Users' Mood from Web Search Query and Mobile Sensor Data, *Big Data*, Vol. 12, No. 3, pp. 191–209 (2024).
- [8] Lee, W.-P., Chen, C.-T., Huang, J.-Y. and Liang, J.-Y.: A smartphone-based activity-aware system for music streaming recommendation, *Knowledge-Based Systems*, Vol. 131, pp. 70–82 (2017).
- [9] Cai, R., Zhang, C., Wang, C., Zhang, L. and Ma, W.-Y.: Musicsense: contextual music recommendation using emotional allocation modeling, *Proceedings of the 15th ACM international conference on Multimedia*, pp. 553–556 (2007).
- [10] Nishiyama, Y., Ferreira, D., Eigen, Y., Sasaki, W., Okoshi, T., Nakazawa, J., Dey, A. K. and Sezaki, K.: IOS crowd-sensing won't hurt a bit!: AWARE framework and sustainable study guideline for iOS platform, *Distributed, Ambient and Pervasive Interactions: 8th International Conference, DAPI 2020, Held as Part of the 22nd HCI International Conference, HCHI 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22*, Springer, pp. 223–243 (2020).
- [11] Nishiyama, Y., Ferreira, D., Sasaki, W., Okoshi, T., Nakazawa, J., Dey, A. K. and Sezaki, K.: Using iOS for inconspicuous data collection: a real-world assessment, *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, pp. 261–266 (2020).
- [12] Hektner, J. M., Schmidt, J. A. and Csikszentmihalyi, M.: *Experience sampling method: Measuring the quality of everyday life*, Sage (2007).
- [13] Pollak, J. P., Adams, P. and Gay, G.: PAM: a photographic affect meter for frequent, in situ measurement of affect, *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 725–734 (2011).