

# 検証レポート - 表情 (smile) 判定

作成者: 藤田光希

日付: 2025-11-20

## 1. 要約

本検証は、与えられた15点の顔ランドマークのみを用いて、軽量かつ実運用向けの「smile / not\_smile 判定ロジック」を設計・評価したものです。

主要な結論は以下の通りです。 -

- 閾値ベースの単純ルールでも約85%の精度を得られる（運用負荷が極めて低い）。
- 機械学習（Logistic Regression / RandomForest）では上限精度が約91.5%に達した。
- 口の「横幅」「縦の開き」「縦横比」が判定に対して最も寄与する特徴である。
- 本ロジックは、アンケート結果や購買データと掛け合わせることで、消費者が「どの広告・商品に最も笑顔反応したか」を軽量に把握でき、店舗導線最適化・広告改善などに展開可能。

## 2. 目的

本作業の目的は以下の通りです。 -

- 15点ランドマークのみで高精度かつ軽量に笑顔判定を行うロジックを検討する。
- 閾値ルールによるベースラインを構築し、その妥当性を検証する。
- MLモデルで性能上限と特徴量の妥当性を確認し、最終的に実運用に適した smile\_predict を設計する。

要件としては「軽量・高速で現場に導入しやすい」ことを優先しています。

## 3. データ構造

各サンプルは次の形式です: [id, p1, p2, ..., p15, {'smile': bool}]。各 pi は (x, y) の座標ペアです。

口周りのランドマーク（本検証で主に使用）: - 12: mouth\_left - 13: mouth\_right - 14: mouth\_top - 15: mouth\_bottom

基礎的な幾何特徴として、以下を計算して評価に使用しました。 - mouth\_width = |x\_right - x\_left| - mouth\_height = |y\_bottom - y\_top| - ratio = mouth\_height / mouth\_width

## 4. 誤り解析 (FP / FN の傾向)

val\_split による誤分類の代表的傾向： - FP:

驚きや発話など口を縦に開けるが笑顔ではないケース。 - FN: 口角だけ上がる small smile (縦の開きが小さく threshold に届かない)。

示唆: - 閾値ルールは mouth\_height / mouth\_width に強く依存するため、small smile を捉えにくい。 - 口角の相対上昇 (left/right corner lift) を閾値に加えることで FN を減らせる可能性が高い。

## 5. 閾値ルール (ベースライン) の検証

### 5.1 方法

val\_split 上で ratio と mouth\_width

に対する閾値をグリッド探索し、最良閾値を選定しました。

探索した候補: - ratios = [0.135, 0.20, 0.21, 0.22, 0.23, 0.24] - widths = [22.5, 30.0, 30.5, 31.0, 31.5, 32.0]

判定ルール: `pred = (ratio > th_ratio) and (mouth_width > th_width)`

### 5.2 結果 (要約)

・ validation で最良だった組合せ: (ratio=0.135, width=22.5) (val accuracy  $\approx$  0.856)

この閾値を各分割に適用した結果:

データセット	Accuracy
train_split	0.859
full_train	0.858
test	0.855

ビジネス的には以下のような利点がある: - 速い (数マイクロ秒) - メモリ不要 - デプロイが最も簡単 (オンデバイスでも動作)

### 5.3 考察

・

単純な2変数ルールで約85%の精度を達成できるため、処理負荷の厳しい環境では十分実用的。

・ 誤分類の傾向: FN (見逃し): 口角だけ上がる small smile (縦高さが小さい)

FP (誤検出): 驚き・発話などで口が大きく縦に開いている非笑顔

・ これらの傾向から、閾値ルールに「口角の上昇」や顔スケール正規化などの追加特徴を組み込むことで改善余地がある。

## 6. 特徴量設計 (Feature Engineering)

精度向上のために `compute_features()` で以下の 11 次元特徴を作成しました (顔スケールで正規化)。

主な特徴: -  $\text{mouth\_openness} = \text{mouth\_h} / \text{face\_scale}$  -  $\text{mouth\_hw\_ratio} = \text{mouth\_h} / \text{mouth\_w}$   
-  $\text{mouth\_w\_over\_face} = \text{mouth\_w} / \text{face\_w}$  -  $\text{mouth\_h\_over\_face\_h} = \text{mouth\_h} / \text{face\_h}$  -  
 $\text{corner\_asym} = |\text{y\_left} - \text{y\_right}| / \text{face\_scale}$  -  $\text{left\_corner\_lift}$ ,  $\text{right\_corner\_lift}$  =  
口角の相対位置 -  $\text{face\_scale}$  = 顔全体のスケール (正規化用)

効果: 顔サイズや撮影距離によるスケール差を吸収し、口角の微小な上昇 (small smile) 検出に寄与します。

## 7. 機械学習モデルによる性能検証

モデルは標準化 (StandardScaler) を行った上で構築しました。

### 7.1 Logistic Regression (LR)

- ・ バリデーションでの簡易 C 探索の結果:  $\text{val accuracy} \approx 0.947$
- ・ テスト精度:  $\text{test accuracy} \approx 0.915$

重要度 (係数) 上位 (概略): -  $\text{mouth\_w\_over\_face}$  (口の横幅) -  
 $\text{mouth\_h\_over\_face\_h}$  (口の高さ) -  $\text{mouth\_openness}$  (開き具合) -  
 $\text{mouth\_hw\_ratio}$  (縦横比)

低寄与の特徴: -  $\text{corner\_asym}$ 、頬・目関連 (今回のデータでは変化が小さい)

### 7.2 RandomForest (RF)

- ・ GridSearchCV (軽量設定) を実施。
- ・  $\text{test accuracy} \approx 0.915$  (LR と同等)

`feature_importances` 上位: -  $\text{mouth\_h\_over\_face\_h}$ ,  $\text{mouth\_openness}$ ,  $\text{mouth\_hw\_ratio}$ ,  
 $\text{mouth\_w\_over\_face}$

### 7.3 総括

- ・ LR と RF が同様の重要特徴を示し、幾何学的特徴が一貫して有効であることを裏付ける。
- ・ ML モデルの上限精度は約 0.915 程度であり、閾値ベース (約 0.85) と比べて改善余地が明確。

## 8. 結論と設計方針 (smile\_predict)

- ・ 結論
- ・ 閾値ベース：実装コスト・実行コストが最小で運用しやすく、精度は約85%。
- ・ ML ベース：上限精度は約91.5%。バッチ学習やオンデバイス推論で高精度化可能。
- ・ 実運用案：軽量条件下：閾値ルールを本番稼働（速く・安価）。高精度が必要な場合：LRを採用し、特徴量に `corner_lift` を含める。両者を組み合わせるハイブリッド運用も検討（閾値で高速フィルターMLで精査）。
- ・ 本判定ロジックは、モデル定義の有無や環境制約に依存せず、「高精度で軽量」の両方を満たす柔軟なハイブリッドアプローチを取る。
- ・ 設計判断の理由：
  - ・ Logistic Regression と RandomForest の test 精度はいずれも約 0.915 と高精度であるが、実運用を考慮すると LR は軽量かつ推論はシンプルな線形計算のみで高速、また重みベースで特徴量の寄与を直接確認でき透明性が高いため、ML ベース smile\_predict を採用する際の第一候補となる。一方、閾値ルールは約 0.85 の精度で極めて軽量であり、モデル・依存パッケージ不要でオンデバイス環境に適する。これらの利点を踏まえ、最終的な smile\_predict は「ML ベースを優先し、fallback として軽量な閾値ルールを組み込む構成」とした。これは、運用環境・精度要件に応じて柔軟に切り替え可能なハイブリッド設計である。

## 9) smile\_predict 関数の実装とテスト

目的： - 実運用環境で利用可能な、軽量・高速かつ堅牢な `smile_predict(points)` を最終的に定義する。本関数は、学習済みモデルを使用可能な環境と、モデルが利用できない軽量環境のどちらにも対応する。

実装方針（優先順位）：

Logistic Regression (`clf_lr`) + StandardScaler (`scaler`) を優先して使用

ML モデルの test accuracy は 約 91.5% と、今回の上限精度を達成。

計算量が小さく、線形モデルのため推論が高速で透明性も高い。

学習済みモデルが利用できない場合は閾値ルールへフォールバック

ランドマーク (12, 13, 14, 15) から `mouth_w`, `mouth_h`, `ratio` を計算。

`val_split` で探索した最良閾値 `best_param = (ratio=0.135, width=22.5)` を使用。

`best_param` が未定義の場合は同じ値をデフォルトとして利用する。

戻り値は True/False の2値

ML が使える場合 : `clf_lr.predict`

ML が使えない場合 : 閾値ルール (`ratio > th_ratio`) and (`mouth_w > th_width`)

テスト実行セル: - ノートブックには `test` データセットに対して上記 `smile_predict` を適用し、正解数・不正解数・精度 (%) を出力する簡易評価セルが追加されています。 - 閾値ルールが `best_param` を使う場合は、レポート中で示した閾値ベースの `test` 精度 (約 85.5%) と一致するはずですが、学習済みの `clf_lr` が存在すれば ML の `test` 精度 (約 91.5%) が出ます。

実装上の注意点と運用示唆: - ノートブックの `smile_predict` は「まず学習済みモデルを使い、無ければ閾値ルールを使う」ハイブリッドな仕組みになっており、実運用での柔軟性が高く、環境依存性が最小化できる。 - ローカル・オンデバイス環境、クラウド API、バッチ処理など 多様な実行環境で安定動作する。 - 例外処理は握りつぶさず、ML が正常に動作しない場合は必ず `fallback` が動くため信頼性が高い。 - 閾値ルールは高速かつ依存ライブラリが不要で、レイテンシ要件が厳しい現場への導入に適する。

## 10) 全体を通した感想・考察

- ・ `mouth_h_over_face_h` が最も効く理由

- ・ 口の縦幅そのものではなく「顔の大きさで正規化した口の縦幅」が効く

`mouth_h_over_face_h = mouth_h / face_height`

は、「顔全体に対する相対的な口の開き」という個人差に依らない指標に変換する。

- ・ このデータセットでは「笑顔＝口が縦に開く」という定義が強い FP/FN

の誤り解析からも明確で、FN＝微笑み（口がほぼ開かない）

-FP＝口が縦に少し開いた非笑顔が中心だった。

- ・ つまり、このデータの `smile` というラベルは「口の縦方向の開き」

に最も強く依存している。`mouth_h_over_face_h` はまさに “口の縦方向の開き” を直接最適化した指標。

- ・ `small smile` が難しい理由

- ・ このデータでは、「口の開き具合」と「顔サイズに対する相対的な口の高さ」がほぼすべての情報を持っており、15点すべてを複雑に使うよりも、口周り＋スケール正規化という少数の幾何特徴だけで 90%超の精度が出る、という構造が本質です。

- ・ 閾値ルール (`mouth_h/mouth_w` と `mouth_width`) だけでも Test

≒85%に達しており、Logistic Regression や RandomForest の導入は「性能の底上げ＋境界の滑らかさ」のための微調整に近い役割になっています。つまり、性能向上は主にモデルの高度化ではなく、「適切な正規化と特徴設計 (`face_scale * mouth_h/face_h` など)」で実現されていると言えます。

- ・ モデル比較から、重要なのは

- ・ 口の縦幅の相対値 (mouth\_h\_over\_face\_h, mouth\_openness, mouth\_hw\_ratio)
  - ・ 口角の位置 (left/right\_corner\_lift) であり、目や頬など他部位の情報はほとんど寄与していません。これは「このデータセットの笑顔定義＝口元中心」であることを示し、ビジネス上も「最低限、口周りを高精度に取ればよい」という要件につながります。
  - ・ 誤り解析では、FN は「口がほとんど開いていない・ごく軽い笑み」が多く、FP は「口がやや開いているがラベリング上は非笑顔」という境界例でした。これは、アルゴリズムの限界と  
 いうより「人間ラベルの曖昧さ（軽い微笑をどこまで smile とするか）」に起因する部分が大きく、そこを詰めてもビジネス上の価値は相対的に小さいことを示唆します。
  - ・ 最終的な設計（軽量の Logistic Regression + スケーラ+閾値ルール of フォールバック）は、
  - ・ 実装・推論コストが非常に低い
  - ・ 特徴の意味が明確で、マーケ／クライアントに説明しやすい
  - ・ モデル更新が容易（閾値や係数の再学習だけ）
- という点で、「リアルタイムなマーケティング調査に組み込みやすい」バランスの良いソリューションになっている、というのが本質的な結論です。
- ・ 口だけのランドマークで smile をやる限界
  - ・ 結果： 本物の笑顔か “作り笑い” かの識別は困難 驚き・発話と small smile の誤判定も発生
  - ・ つまり「口だけのランドマーク」は笑顔分類の
- \*\*80%程度は説明可能だが、完全ではない」という明確な限界がある。

（終）