

構造化LLMセマンティクスと2次元YIN/YANG理論による 全体的なファッション分類

Minami YODA
Nihon University
Tokyo, Japan
yoda.minami@nihon-u.ac.jp

Jialong Li
Waseda University
Tokyo, Japan
lijialong@aoni.waseda.jp

Yasuyuki TAHARA
The University of Electro-Communications
Tokyo, Japan
tahara@uec.ac.jp

Yuichi SEI
The University of Electro-Communications
Tokyo, Japan
seiuny@uec.ac.jp

松野豊
Nihon University
Tokyo, Japan
松野・湯中@ニホン-u.ac.jp

要旨-パーソナルカラー分析やボディフレーム分析を含む画像コンサルティング(IC)は、スタイリングや購買の意思決定を支援するために広く利用されている。AIベースのICは、色解析や体型推定などの要素レベルのタスクのために研究されてきたが、人物の全体的な外観の全体的な分類はまだ十分に研究されていない。専門的な実践において、YIN/YANG理論は、このような全体的な診断のための一つの枠組みを提供し、顔の構造、ボディフレーム、ラインの品質、知覚される成熟度を統合して、調和するファッション要素を導き出す。しかし、専門家による診断の再現は依然として困難である。我々の予備調査では、同じ個人の診断は約60%の症例で分析者間で異なることが示された。

この問題に対処するため、我々はYIN/YANG理論を従来の一次元軸から、線の質と知覚される成熟度からなる二次元空間に再定式化する。さらに、大規模言語モデル(LLM)によって生成された構造化記述を中間表現として用いる、全体的な分類フレームワークを提案する。このモデルは、顔の特徴、骨の構造、体長、曲率、知覚される成熟度を記述するよう促され、これらの記述は、分類のために解釈可能な数値特徴に変換される。100人の有名人を含むベンチマーク評価において、GPT-5-miniとkNN(k = 1)分類を用いた構成は、55.00%のTop1精度と56.44%のWeighted F1スコアを達成した。これらの結果は、構造化されたLLM記述が、再現可能なYIN/YANGベースの分類をサポートできることを示唆しており、高度に主観的な診断タスクのセカンドオピニオンとして機能する可能性がある。

索引用語-ファッションAI、大規模言語モデル、YIN/YANG理論、ファッション分類、

I. はじめに

個人の色彩分析やボディフレーム分析を含む画像コンサルティング(IC)は、スタイリングや購買の意思決定をサポートする[1]。ICは、人々が自分の身体的属性により適合した衣服を選択できるようにすることで、より長い使用とより持続可能な消費を促すこともできる[2]。

AIベースのICは、個人の色分析、体型推定、衣服の推奨のために研究されているが[1]、[3]~[8]、ほとんどの研究は、人物の全体的な評価ではなく、個々の要素に焦点を当てている。視覚言語モデルを用いた互換性評価

が報告されているが[9]、理論に基づく全体的な分類はまだ十分に検討されていない。

プロフェッショナル・スタイリングは、複数の物理的特徴を統合する必要があり、YIN/YANG理論は、そのような属性からファッション要素を調和させるフレームワークを提供する[10]、[11]。しかし、これらの特徴は相互に依存しているため、全体的な診断を再現することは困難である。239の公開投稿から57の重複診断を調査したところ、約60%が分析者の間で意見が一致しなかった。この矛盾は、客観的で再現性のある計算手法の動機付けとなる。

このギャップを埋めるために、本研究では、構造化LLM生成セマンティクスを用いてYIN/YANG理論を形式化した、説明可能な最初のAIフレームワークを提案する。元の1次元分類軸を2次元空間に拡張し、顔の特徴、骨格構造、知覚される成熟度に関するLLM記述を分類のための数値特徴に変換する。

本研究の貢献は以下の通りである：

- 我々は、全体的な外観分類における分析者間の一致を分析し、人間の診断の不安定性を強調した。
- YIN/YANG理論を2次元空間に再定式化し、構造化記述とルールベースの特徴構築を組み合わせた解釈可能なAIフレームワークを開発した。
- 100人の有名人のデータセットを用いて、トップ1の精度を55%、重み付けF1スコアを56.44とし、人間変動が大きい診断タスクに対するAIベースの自動化の可能性を示した。

II. 関連研究

A. 分類と推奨事項

AIベースのICに関する先行研究では、個人の色分析、体型推定、衣服の推薦、ファッション画像-テキスト表現学習が扱われている[1]、[3]~[8]。これらの研究は、人間や衣服の属性をスタイリングサポートに結びつけているが、

TABLE I
INTER-ANALYST DIAGNOSTIC RESULTS BASED ON PUBLIC DIAGNOSIS DATA

Category	Count	Ratio
239件の公開投稿の内訳		
Reference analyst only	65	27.1%
Related analyst only	117	48.9%
両方により診断される	57	23.8%
57例の二重診断の内訳		
Exact agreement	9	15.8%
主な分類の不一致	35	61.4%
サブ分類の不一致のみ	13	22.8%

主に肌の色、体型、衣服の属性など、個々の要素を対象としている。VLMに基づく人と着用衣服の適合性評価が報告されているが[9]、理論に基づく人全体の全体的な分類はまだ限定的である。

B. YIN/YANG理論

YIN/YANG理論は、人の身体的特徴や印象を直線性、曲率、関連属性の観点から特徴付け、それらと調和するファッション要素を導き出すためのフレームワークである[10], [11]。有名人のスタイリングなど、専門的な文脈でも利用されている。元の理論には6つのタイプがある：ドラマチック、ナチュラル、クラシック、ロマンティック、インジーン、ガミンである。日本のYIN/YANG理論では、主分類と副分類が割り当てられている。主な分類は、適切なファッション要素の広範な範囲を決定し、副分類は、より細かい差別化を提供し、よりパーソナライズされたレコメンデーションを可能にする。

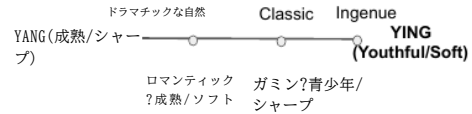
C. 拡張YIN/YANG理論

図1にYIN/YANG理論の分類マップを示す。元の理論は、直線から曲線への一次元連続体である。しかし、この表現では、ロマンティック(曲線/成熟)やガミン(直線/若者)のように、線の質と知覚される成熟度が交差するタイプを適切に扱うことができない。この限界に対処するために、現代のIC診療では2次元的な解釈が一般的になっている。

III. Sアナリスト間の診断の一貫性に関する調査

拡張YIN/YANG理論に基づく全体分類法の診断における分析者間の整合性を検討した。Xとノートに掲載された239件の一般公開された診断投稿を収集し、同一人物が日本の大手診断医とその他の診断医の両方から診断を受けた症例を抽出した。結果を表Iに示す。比較可能な57症例があり、それらは3つのカテゴリーに分類された：厳密な一致、主分類における不一致、副分類における不一致のみ。厳密な一致率はわずか15.8%であったが、61.4%の症例が主分類レベルで不一致であった。これらの結果は、ヒトの診断結果が不安定であることを示唆している。

元の1次元モデル



拡張2次元モデル

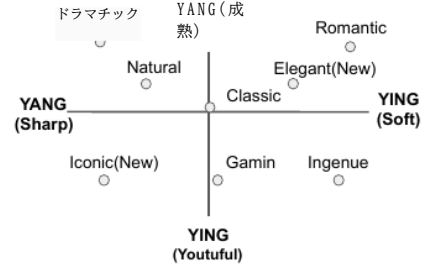


図1. オリジナルモデルと拡張モデルの分類マップ



図2. 拡張YIN/YANG理論から生成された8種類の女性像

IV. 提案された方法

A. 拡張YIN/YANG理論の定式化

上記の2次元解釈に基づき、我々は拡張YIN/YANG理論を、直線/YANGから曲線/YINへの線の質と、成熟/YANGから若年/YINへの成熟の2軸を持つ計算特徴空間として定式化する。この定式化は、ラインの品質と知覚された成熟度を分離し、元の理論では1つの軸に沿って混合された。元の6つのタイプがこの2次元空間にマッピングされると、古典派とローマ字派の間に、また若々しい/まっすぐな側に、カバーされていない領域が現れる。これらの領域をカバーするために、エレガントとアイコンックを導入し、8つのラベルの分類法を採用する：エレガントは古典とローマ字の間の曲線/成熟領域を表し、アイコンックはガミンよりも若々しい/直線領域を表す。図2は、YIN/YANG理論を2次元に拡張して生成された8種類の女性画像である。

B. ワークフロー

提案するワークフローでは、まず各人の複数の画像がLLMに入力され、顔の形状、輪郭、骨構造、ボディフレーム、曲率、知覚される成熟度、キーワードなど、あらかじめ定義されたスキーマに従って構造化された記述を生成する。これらの記述は、次に数値特徴に変換され、人物レベルのベクトルに平均化され、k-最近傍を使用して分類される。

C. プロンプト設計

本手法は、人物の外見印象を構成する安定した物理的構造の記述を生成する。この設計では、LLM出力をブラックボックス的な最終決定としてではなく、分類器の中間表現として使用する。プロンプトは、顔の形、輪郭、顎、頬骨、目の形状、鼻と口の構造、骨格、体長、曲率、知覚される成熟度を観察するようモデルに求める。一方、衣服、ヘアスタイル、化粧、ポーズ、背景、表情、気分や性格の推測、有名人のアイデンティティ、過去の一般人のイメージ、実際の年齢推定は明示的に除外されている。

D. 特徴量の定量化と分類

構造化されたLLM記述を得た後、ルールベースの語彙マッピングにより数値特徴に変換する。各画像は11次元の特徴ベクトルで表現される:6つの顔形状特徴、2つの線質特徴、2つの知覚年齢特徴、および補助バランス特徴からなる10次元の基本ベクトルである。したがって、記述は、分類ラベルに直接マッピングされるのではなく、解釈可能な中間特徴にマッピングされる。

顔形状の特徴は輪郭の形状を表し、線の質は線がどれだけ湾曲しているか、柔らかいか、強いかを記述する。知覚年齢特徴量は、実際の年齢ではなく、外見構造に由来する成熟した若々しい印象を表し、補助的なバランス特徴量は、顔の特徴配置における対称性または中立的な位置の記述に対応する。

w_s をスキーマフィールドsの重みとし、画像iのフィールドsが特徴fに関連する項を含むかどうかを示す非負のマッチングスコアを $m_{i,s,f}$ とする。画像レベルの特徴量 $x_{i,f}$ は以下のように計算される:

$$x_{i,f} = \frac{\sum_s w_s m_{i,s,f}}{\sum_s w_s}.$$

例えば、丸い輪郭の記述は丸みや曲率に関連する特徴を、直線的な記述は直線性に関連する特徴を、成熟や洗練などの表現は成熟に関連する特徴を活性化する。分類器として、軽量で解釈可能なk-nearest neighbors法を用いる。k = 1 のとき、特徴空間における最も近い訓練人物のラベルが直接予測として使用される。

TABLE II
タイプラベルの配布

Type Label	Count
Romantic	26
Classic	22
Natural	25
Elegant	2
Ingenue	16
Dramatic	9
Total	100

V. 実験と結果

A. データセットと構成の最適化

実験には、YIN/YANGの分類タイプが既知の有名人100人を使用し、個人ごとに最大100枚の画像を収集した。本評価では、表IIに示すように、主な分類ラベルのみを使用した。データセットは、トレーニング用に80個体、評価用に20個体に分割された。

評価指標として、Top-1精度、Top-3精度、Macro F1、Weighted F1を用いた。マクロF1はタイプ間でF1スコアを等しく平均化するが、重み付けF1はサンプル数でタイプごとに重み付けを行う。

本実験の前に、LLM間で構造化記述の精度を比較した。結果を表IIIに示す。比較ではサンプルサイズが小さく(n = 9)、評価値が11.11ポイント刻みで変化したため、モデルの優位性に関する決定的な結論を導き出すことは困難である。ただし、コストと精度のバランスを考慮し、GPT-5-miniを採用した。

次に、10枚、20枚、30枚、40枚、50枚の画像を用いて画像カウントスイープを行い、人物レベルの特徴に必要な画像数を求めた。表IVに示すように、性能は10画像から20画像に向上したが、その後は単調に向上しなかった。したがって、本実験ではGPT-5-miniを使用し、1人当たり20枚の画像を作成した。

本実験では、10次元の特徴ベクトルであるshape_curve_age (SCA)と、SCAに補助的なバランス特徴を追加する11次元の特徴ベクトルであるshape_curve_age_balance (SCAB)を比較した。バランスは、クラシックのような中程度またはよく比例した印象を捉えることができるが、2次元のYIN/YANG空間の主軸、すなわちストレート/カーブ、成熟/ユースに直接対応しないため、これらの設定を別々に評価した。そこで、本研究では、バランス機能を用いないSCAとバランス機能付きSCABを比較し、バランス機能の有用性を検討する。

B. 主な結果

主な実験結果を表Vに示す。最良の構成はSCA + kNN-1で、Top1精度55.00%、Weighted F1スコア56.44を達成した。SCA + kNN-3は70.00%のトップ3精度を達成したが、バランスを加えたSCABはSCAを上回らなかった。このことから、今回の実験条件では、

TABLE III
LLMスクリーニングの結果(MALL-Scale条件下) (n = 9)

Model	Top-1	Top-3	W-F1
GPT-5	44.44	66.67	27.35
GPT-5-mini	33.33	77.78	26.67
GPT-5-nano	44.44	66.67	32.32
GPT-4o	33.33	77.78	26.67
Claude 3.5 Sonnet	44.44	66.67	27.35
Gemini Flash	33.33	77.78	22.22

TABLE IV
1人レベルの特徴を構築するために使用される画像数の影響 (GPT-5Mini)

Images	Top-1	Top-3	Macro F1	W-F1
10	16.16	67.68	11.35	15.86
20	26.26	73.74	18.06	25.48
30	24.24	72.73	16.82	23.50
40	20.20	75.76	13.65	19.27
50	25.25	75.76	17.59	23.83

身体に関連する主な特徴に基づくSCAがより一貫して機能することが示唆された。

VI. ディスカッション

提案手法は、構造化されたLLM記述を数値特徴に変換し、kNNを用いて人物レベルベクトルを分類するため、予測に用いる中間特徴を検査することができる。このように、この方法は、人間の診断のブラックボックス的な置き換えとしてではなく、よりチェック可能な形で診断をサポートする第二の意見として意図されている。同じプロトコルでの人間による評価はまだ行われていないが、55.00%というトップ1の精度は、構造化記述に基づく分類が人間の判断の性能範囲に近づく可能性を示唆している。

A. 制限事項

本研究にはいくつかの限界がある。主な評価は、GPT-5-miniを用いた100人、20画像、固定分割の設定に限定され、他のモデルや100画像までの画像カウンスイープとの比較は未完成のままである。さらに、構造化された記述から数値的特徴への変換はルールベースであり、語彙のバリエーションやプロンプトの設計に影響される可能性がある。

VII. おわりに

本研究では、YIN/YANG理論に基づく外観印象分類を、分類器が利用できる特徴表現として整理し、画像から分類結果を得るための手法を提案した。まず、人間の分析者間の診断の大きなばらつきを調査し、従来の二次元YIN/YANG軸を、直線/曲線、成熟/有色人種の二次元空間として整理した。次に、LLMが顔の形状、輪郭、骨の構造、曲率、知覚される成熟度などの特徴の構造化された記述を出力する

TABLE V
M100人の個人の固定された分割に関する主な結果

コンフィギュレーション	Top-1	Top-3	Macro F1	W-F1
SCA + kNN-1	55.00	55.00	46.21	56.44
SCA + kNN-3	50.00	70.00	31.40	49.30
SCAB + kNN-1	50.00	50.00	40.76	53.31
SCAB + kNN-3	45.00	55.00	26.39	44.56

SCA:形状曲線年齢、SCAB:形状曲線年齢バランス

フレームワークを構築し、これらの記述を分類のための数値特徴に変換する。

100人の有名人を使った評価では、GPT-5-miniベースの構造化記述とSCA + kNN1の組み合わせは、55.00%のトップ1精度と56.44の重み付きF1スコアを達成した。この結果は、LLMに直接クラスラベルを出力するよう求めるのではなく、まずLLMを使って観察可能な特徴を記述することで、分類プロセスを検査しやすくできることを示している。したがって、本研究は、提案手法を、再現性の低い全体的な分類タスクにおいて、人間の判断をサポートする第二の意見として使用する可能性を示唆している。

参考文献

- [1] H. ホン・R・I・キム「個人色彩分析のためのモバイルアプリケーション」コジェント・ビジネス&マネジメント、第6巻、第1号、p.1576828, 2019. [オンライン]. <https://doi.org/10.1080/23311975.2019.1576828>
- [2] C. リチャードソン、R. ボードマン、S. ギル、「生涯延長と循環型廃棄の動機づけとしての衣服の適合性の探求」、リソース、保全、リサイクル、第204巻、p.107494, 2024. [オンライン]. <https://www.sciencedirect.com/science/article/pii/S0921344924000880>
- [3] D. I. ベレット、R. スプレングルマイヤー、「衣服の美学：フェアとタンニングの肌色を一致させるための一貫した色の選択」、i-Perception、12巻、6号、pp.2021年1-22
- [4] S. C. ヒダヤティ、C.-C. Hsu, Y.-T. チャン、K.-L. 華、J. フー、W. H. Cheng, "What dress fits me best? (1)衣服のスタイルがパーソナルボディの形状にどのように適合するか," 第26回ACMマルチメディア国際会議予稿集. 米国ニューヨーク州ニューヨーク:計算機学会、2018年、pp.438-446.
- [5] W. ゴン、F. ベイター、「体型に基づくドレス推奨スタイル嗜好」、国際繊維・アパレル協会年次大会講演論文集、80巻、1号、2024年。
- [6] W.-H. チェン、Y. ソン、C.-H. チェン、S.-F. ヒダヤティ、J. リュー、「ファッションとコンピュータビジョンの出会い: 調査」ACM Computing Surveys, vol. 54, no. 4, pp.2021年1-41
- [7] Z. 劉、P. 羅、S. 秋、X. 王、X. 唐、「Deepfashion: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp.1096-1104
- [8] P. このような、言語と視覚の対比学習は、科学報告書、12巻、1号、p.18958, 2022.
- [9] Y. 平川陽一郎、和田哲也、森下和彦、清水亮、古澤哲也、S. H. Kham, 斉藤恭一, "ファッション美的評価におけるgpt-4vの性能の実証分析," SIGGRAPH Asia 2024 Technical Communications, ser.SA '24. 米国ニューヨーク州ニューヨーク:計算機学会、2024年。[オンライン]. <https://doi.org/10.1145/3681758.3698022>
- [10] B. ノースラップ「コスチュームとパーソナリティの問題へのアプローチ」『アート・エデュケーション・トゥデイ』vol.2, pp.94-104, 1936.
- [11] H. マクジムゼー、衣服選択における芸術とファッション。Iowa State University Press, 1973.