

研究まとめ

診断スコアから **MBTI** を予測することの限界
および潜在的クラスタの考察(仮)

田中直哉

1 研究背景および目的

1.1 データセットについて

Kaggle より引用(Google 翻訳を使用)

説明

この合成データセットは、人口統計学的要因、興味分野、および性格スコアの組み合わせに基づいて、マイヤーズ・ブリッグス・タイプ指標 (MBTI) の性格タイプを探索および予測するために設計されている。10 万以上のサンプルが含まれており、それぞれが MBTI タイプを決定する様々な特徴を持つ個人を表している。このデータセットは、異なる性格特性と、年齢、性別、教育、興味などの外的要因との相関関係を研究するために使用できる。

データセットの説明

Age:個人の年齢を表す連続変数。

Gender:個人の性別を示すカテゴリ変数。可能な値は「男性」と「女性」。

Education:バイナリ変数。値が 1 の場合、個人は少なくとも大学院レベル（またはそれ以上）の教育を受けていることを示し、値が 0 の場合、大学、高校レベル、または無学であることを示す。

Interest:個人の主な興味分野を表すカテゴリ変数。

Introversion Score : 0 から 10 までの連続変数で、個人の内向性または外向性の傾向を表します。スコアが高いほど、外向性の傾向が強いことを示す。

Sensing Score : 0 から 10 までの連続変数で、感覚と直観のどちらを重視するかを表す。スコアが高いほど、感覚が重視されることを示す。

Thinking Score : 0 から 10 までの連続変数で、思考と感情のどちらを優先するかを示す。スコアが高いほど、思考を優先する傾向がある。

Judging Score : 0 から 10 までの連続変数で、知覚よりも判断を好む傾向を表す。スコアが高いほど、判断を好む傾向を示す。

Personality:人物パーソナリティタイプを含むターゲット

引用終わり [1]

表 1-1 データセット最初の 5 行

Age	Gender	Education	Introversion Score	Sensing Score	Thinking Score	Judging Score	Interest	Personality
21.0	Female	1	5.89208	2.144395	7.32363	5.462224	Arts	ENTP
24.0	Female	1	2.48366	3.206188	8.06876	3.765012	Unknown	INTP
26.0	Female	1	7.02910	6.469302	4.16472	5.454442	Others	ESFP
30.0	Male	0	5.46525	4.179244	2.82487	5.080477	Sports	ENFJ
31.0	Female	0	3.59804	6.189259	5.31347	3.677984	Others	ISFP

1.2 研究背景 (MBTI の特徴量) および目的と課題

現在 MBTI の応用研究は Google Scholar で 2 万 1100 件[2]、CiNii で 94 件[3]、PubMed では 234 件がヒットし年々上昇傾向で[4]、JSTOR では 437 件[5]と少なくない研究論文が存在し、現在も MBTI を用いた研究が進んでいる。

また、MBTI の活用例ではオンライン授業の好み[6]や会社内の最適配置などの研究がある[7]。MBTI は 93 個の質問によって構成されており、その精度が 100%と仮定して、実際に 93 個の質問に答えることは質問者にとって敷居が高く回答者は 93 個の質問を五段階で答えるため曖昧になり解答疲れや適当な回答をする可能性が否定できない。実際に精度に限界性について指摘されており[8]、短縮版での精度の不一致率は 20%から 35%程度になった研究もある[9]。

そこで本研究では逆に内向スコア・感覚スコア・思考スコア・判断スコアと回答者の属性として年齢・性別・教育・興味の 8 個の変数を用いて実際に精度を逆から測定する。ここで精度が 93 個の質問と同等であれば 8 個の変数で代替できるが、精度が落ちてしまった場合はどういう間違え方をしていて、どの特徴量が影響していたか検証する必要がある。

2 研究手法

2.1 精度の測定 (検証)

クロスバリデーションで 500 分割して精度の検証を行う。

ここではクロスバリデーションの結果を要約統計量で平均値と中央値を計測し、ヒストグラムを出力することで最頻値(仮)を確認する。

2.2 分類寄与率の測定

GBDT は決定木系アルゴリズムであるため一つの変数でどれだけ分類できるか測定することができる。そこで、各変数が 2.1 項で担保された精度で MBTI を分類するうえでの影響度を測定する。

また、ハイパーパラメータは LightGBM[10]の初期値を使用する。

2.3 間違った傾向についての分析

Hold-out 方式を用いて訓練データ 70%でテストデータ 30%とし、2.1 項において精度の代表値に近い精度を出したときに Classification_Report(各目的変数の値毎に)と混合行列を出力し、誤った傾向について考察する。

2.4 MBTI 別特徴量のヒストグラム

MBTI 別に特徴量をヒストグラムで画像を保存する。この画像を MBTI 別に比較することで MBTI 別の特徴量の傾向を分析する。

3 結果

3.1 精度の測定

クロスバリデーションで 500 回精度の検証を行った結果得られたようやく統計量を表 3-1 に記す。また、ヒストグラムを図 3-1 に示す。

表 3-1 500 回クロスバリデーションした結果得られたようやく統計量

cross_validation	
count	500.000000
mean	0.895525
std	0.032741
min	0.806818
25%	0.873563
50%	0.896552
75%	0.919540
max	0.988506

表 3-1 の結果から平均値および中央値ともに約 89%という事が分かる。

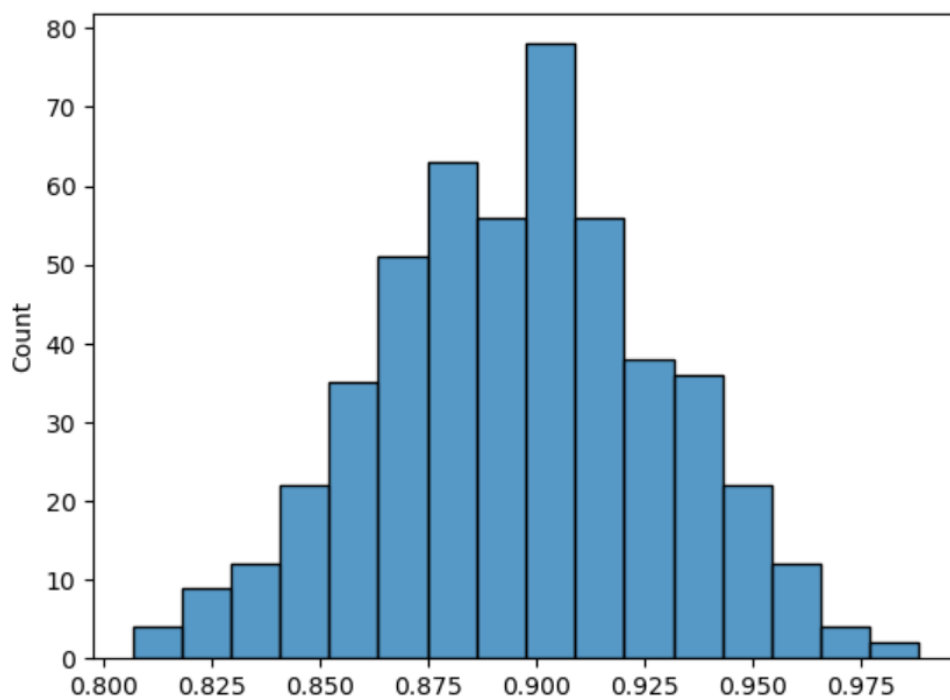


図 3-1 500 回 CV を行った際に得られた精度のヒストグラム(最頻値は 90%)

3.2 分類寄与率の測定

GBDT で MBTI における各特徴量について 89%の信頼度で分類寄与率を測定した結果を表 3-2 に記す。

表 3-2 GBDT を用いた分類寄与率(信頼度 90%)

	imp
Thinking Score	0.199917
Judging Score	0.196146
Introversion Score	0.186438
Sensing Score	0.175604
Age	0.105313
Gender_Male	0.034042
Education	0.032250
Interest_Arts	0.023229
Interest_Sports	0.020646
Interest_Unknown	0.011729
Interest_Technology	0.008146
Interest_Others	0.006542

この結果から各種スコアと年齢が MBTI に関係していることが考えられる。
一方で教育や性別、興味などは MBTI の分類結果に寄与しにくいことも考えられる。

3.3 間違っただ傾向についての分析

精度が代表値だった時の Classification_Report を図 3-2 に、混合行列を表 3-3 に記す。

	precision	recall	f1-score	support
ISFP	0.83	0.91	0.87	278
ISTP	0.89	0.94	0.91	277
INTP	0.85	0.88	0.87	247
ENTP	0.92	0.85	0.88	284
INFJ	0.91	0.87	0.89	264
ESTP	0.90	0.90	0.90	272
ESFP	0.84	0.87	0.86	259
ENFJ	0.92	0.89	0.91	267
INFP	0.90	0.86	0.88	309
INTJ	0.92	0.87	0.89	290
ISTJ	0.85	0.97	0.91	261
ENFP	0.90	0.85	0.87	269
ESFJ	0.90	0.88	0.89	293
ISFJ	0.84	0.90	0.87	284
ENTJ	0.91	0.88	0.90	249
ESTJ	0.95	0.90	0.92	272
accuracy			0.89	4375
macro avg	0.89	0.89	0.89	4375
weighted avg	0.89	0.89	0.89	4375

図 3-2 MBTI を特徴量で分類予測した時の Classification_Report

f1 スコアに着目すると、どの MBTI も正解率の代表値程度(誤差±3 程度)に間違えていることが分かる。

表 3-3 MBTI を分類予測した際の混合行列

	ISFP	ISTP	INTP	ENTP	INFJ	ESTP	ESFP	ENFJ	INFP	INTJ	ISTJ	ENFP	ESFJ	ISFJ	ENTJ	ESTJ
ISFP	252	0	0	0	0	0	22	0	4	0	0	0	0	0	0	0
ISTP	1	259	1	0	0	15	0	0	0	0	1	0	0	0	0	0
INTP	0	12	217	15	0	0	0	0	2	1	0	0	0	0	0	0
ENTP	0	0	32	240	0	12	0	0	0	0	0	0	0	0	0	0
INFJ	0	0	0	0	230	0	0	18	1	1	0	0	0	14	0	0
ESTP	0	20	0	5	0	246	1	0	0	0	0	0	0	0	0	0
ESFP	31	0	0	0	0	0	225	0	0	0	0	3	0	0	0	0
ENFJ	0	0	0	0	16	0	0	238	0	1	0	0	9	2	1	0
INFP	17	0	1	0	0	0	1	0	267	0	0	23	0	0	0	0
INTJ	0	0	3	0	1	0	0	0	0	252	15	0	1	0	18	0
ISTJ	0	0	0	0	0	0	0	0	0	1	254	0	0	0	0	6
ENFP	0	0	0	2	0	0	17	0	22	0	0	228	0	0	0	0
ESFJ	0	0	0	0	0	0	1	1	0	0	0	0	258	33	0	0
ISFJ	1	0	0	0	5	0	0	1	0	0	1	0	20	256	0	0
ENTJ	0	0	0	0	0	0	0	0	0	19	2	0	0	0	220	8
ESTJ	0	0	0	0	0	0	0	0	0	0	25	0	0	0	2	245

誤分類傾向としては、まず I と E(内向的か外向的か)で間違えており、その次に N(直感型)と S(感覚型)で間違えている。

間違えている項目を見てみると I と E は多くて正解に対して 10%以上間違えており、S と N は誤分類場合が多く、5%程度間違えている(図 3-3)。

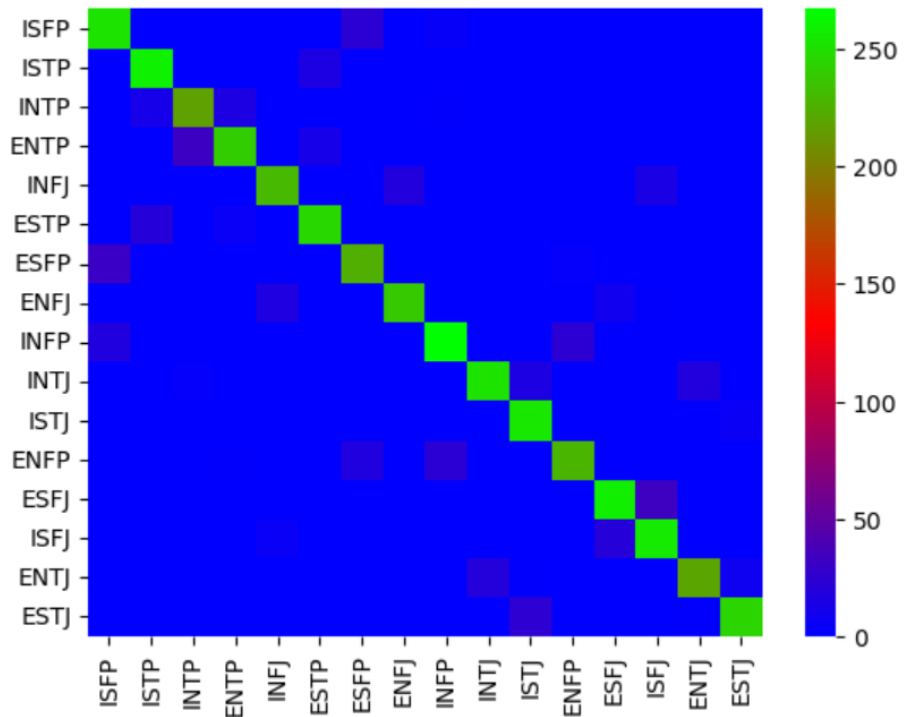


図 3-3 MBTI 分類における混合行列の列のヒートマップ

3.4 MBTI 別特徴量のヒストグラム

MBTI 別に各特徴量をヒストグラムにした結果を図 3-4 から図 3-11 まで示す。

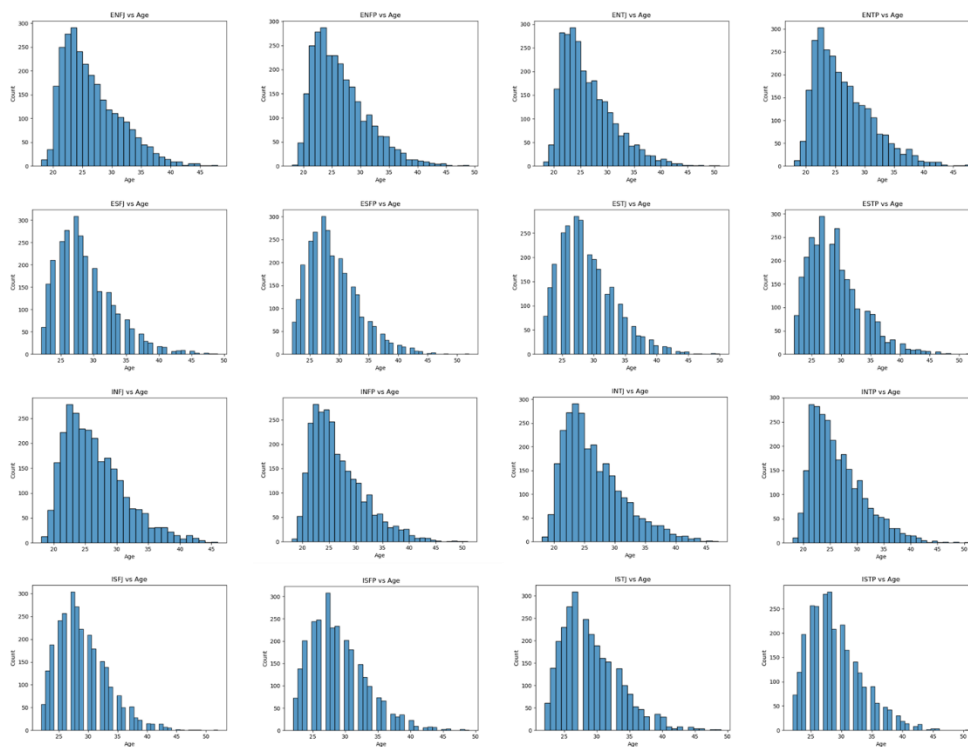


図 3-4 MBTI 別年齢

図 3-4 から MBTI において年齢による差は少ないが、分布から若年層を中心としたデータセットであることが分かる。

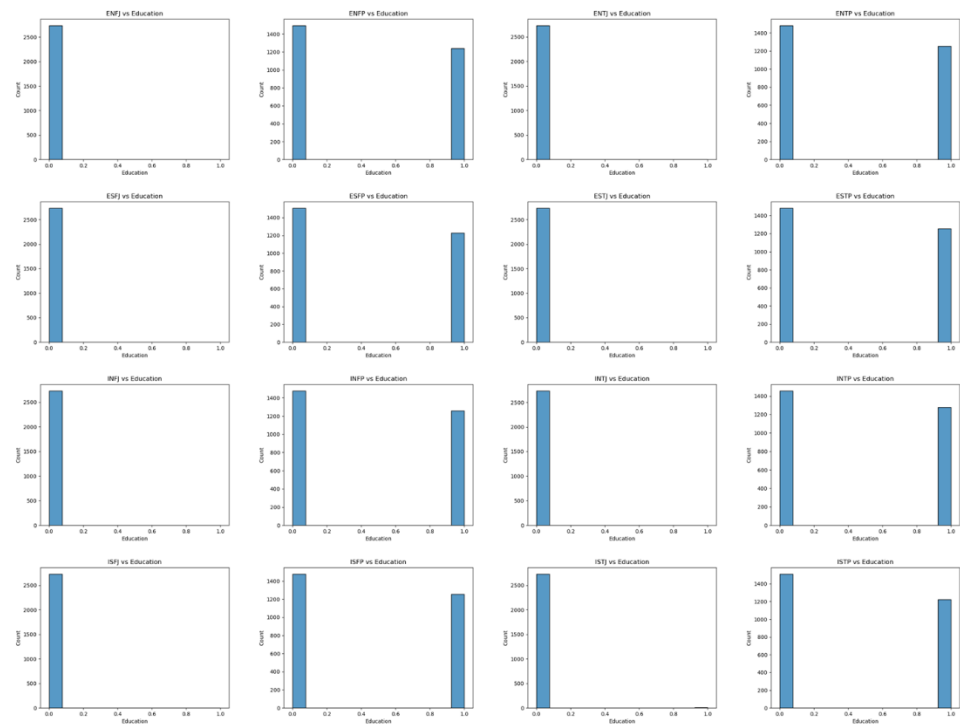


図 3-5 MBTI 別教育

図 3-5 から本データセットにおいて J(判断型)と P(知覚型)では知覚型の P は大学院進学あるいはそれ以上の結果になり J は大学進学までという結果になった。

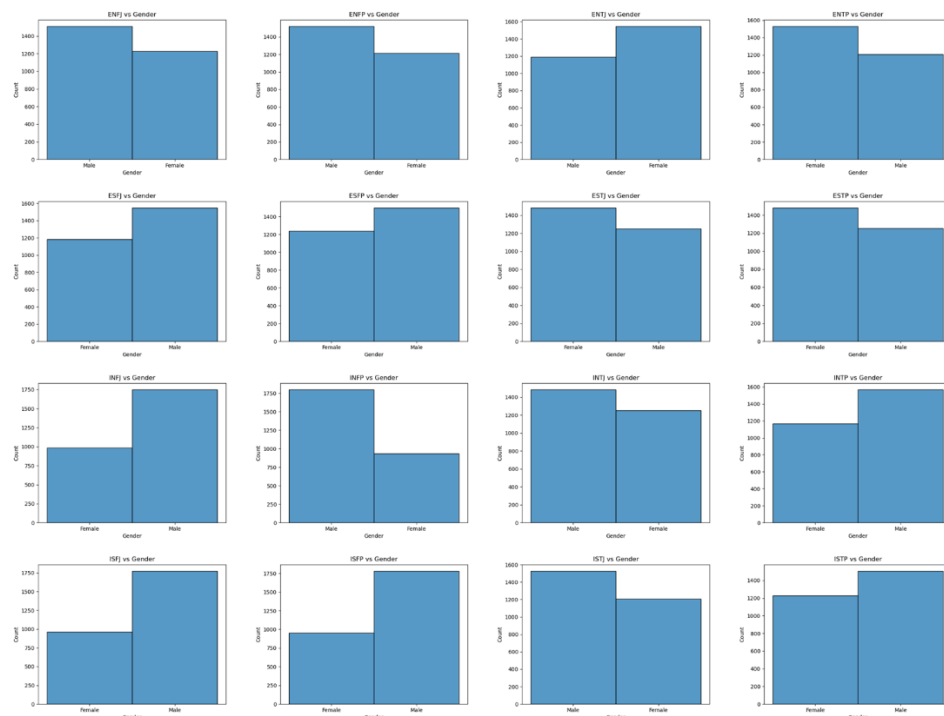


図 3-6 MBTI 別性別

図 3-6 より E(外向的)の性別は S(感覚型)か N(直感的)かで男女比が逆転しており TJ(論理的思考)の場合、他の S か N と比率が逆転している。また、N の時は男性が多く、S の時は女性が多い。

また、I(内向的)の性別は INFJ と ISFJ で同じ性別の多さは同じ程度の比率だが E と比べて男女比が大きく異なる。同様に、INFP と ISFP は性別が逆転しているが E と比べて男女比が大きく異なる。

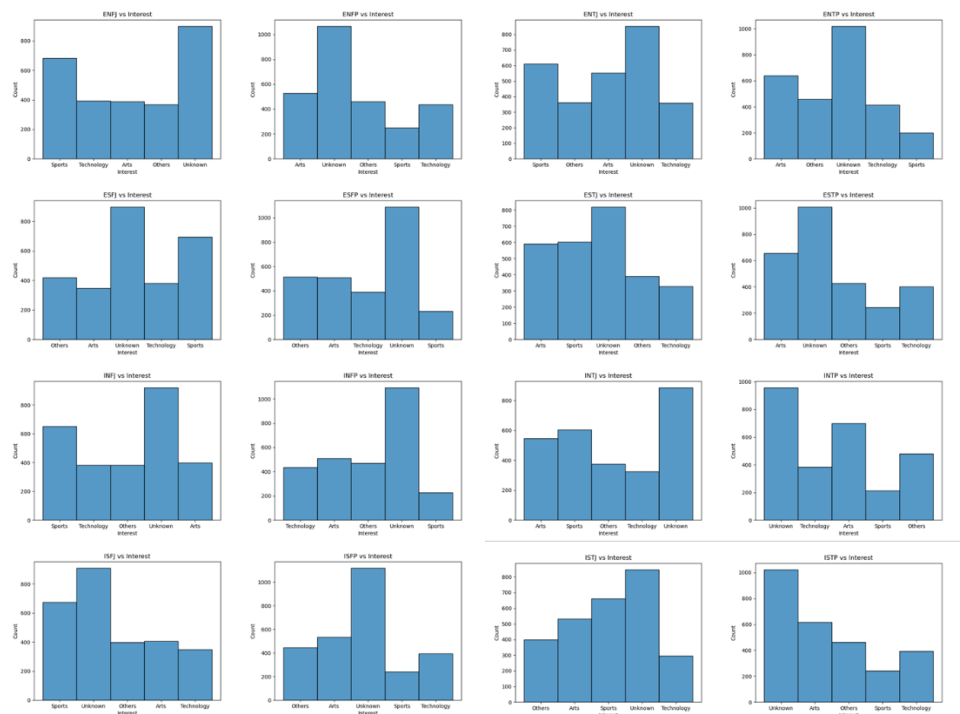


図 3-7 MBTI 別興味

図 3-7 より明確な傾向性自体は確認できなかった。ただし、MBTI 別に何に興味があるかは分かれていることが分かった。

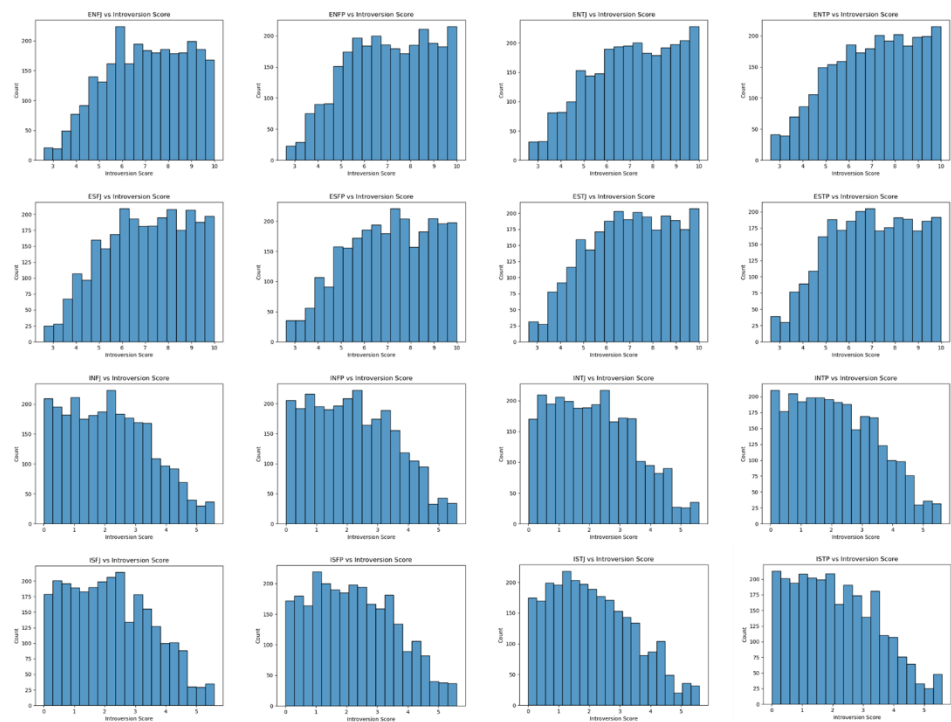


図 3-8 MBTI 別内向スコア

図 3-8 から基本的に E(外向的)は右肩上がりの結果だったのに対して、I(内向的)は右肩下がりの結果になった。

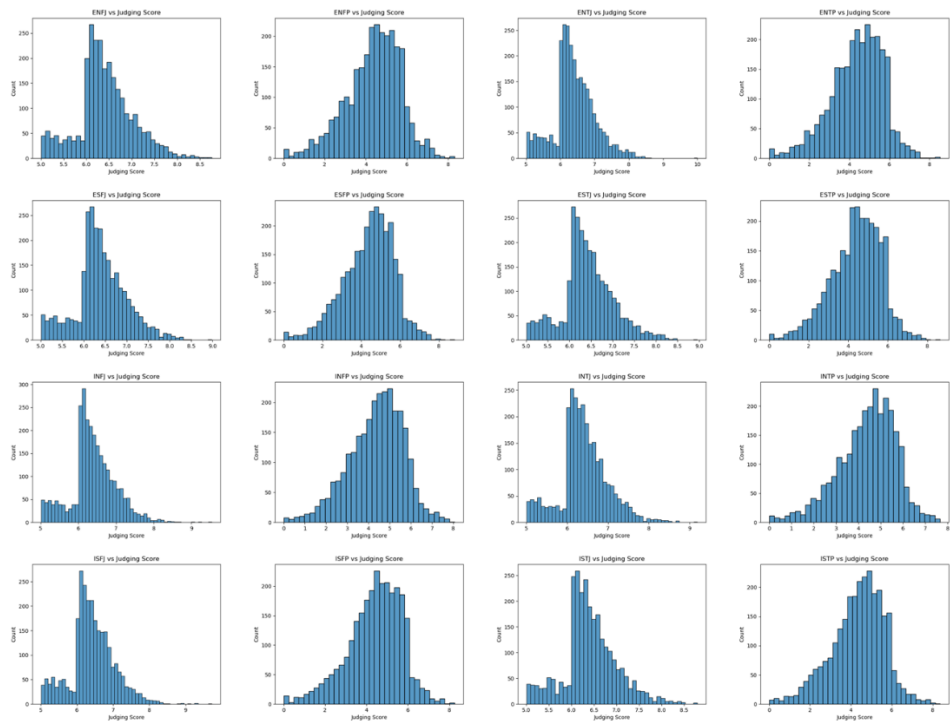


図 3-9 MBTI 別判断スコア

図 3-9 より J(判断型)はスコアが 6 辺りから急激に値が上昇しているおり、その後すぐ下降している。それに対して P(知覚型)はどれも近似した分布をしていることが分かる。

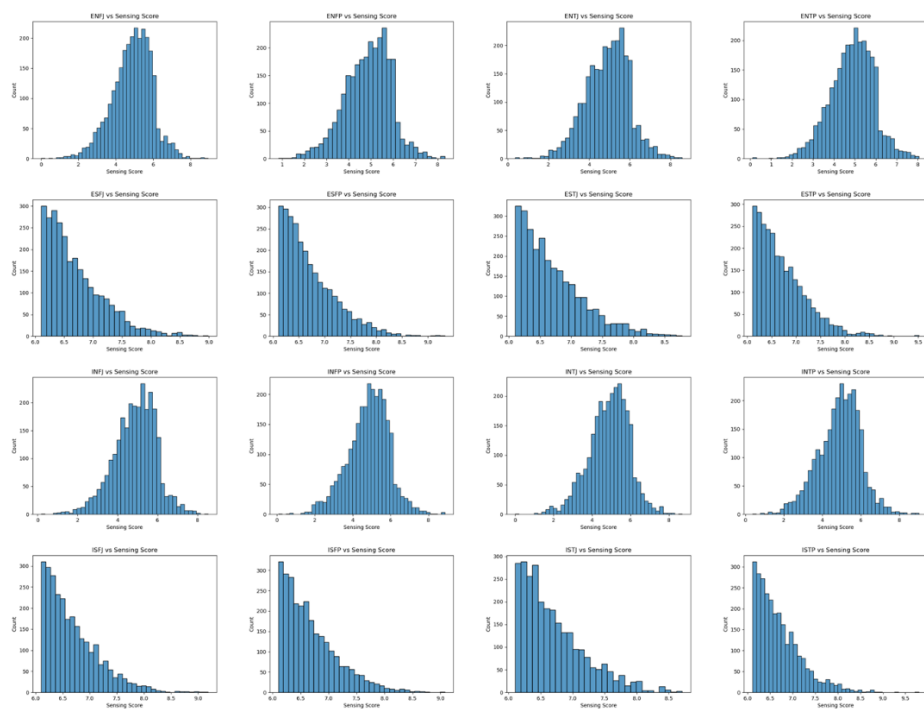


図 3-10 MBTI 別感覚スコア

図 3-10 から S(感覚型)はスコア 6 近くに最頻値がありそこから下降している。
一方で N(直感型)は基本的に似た分布をしている。

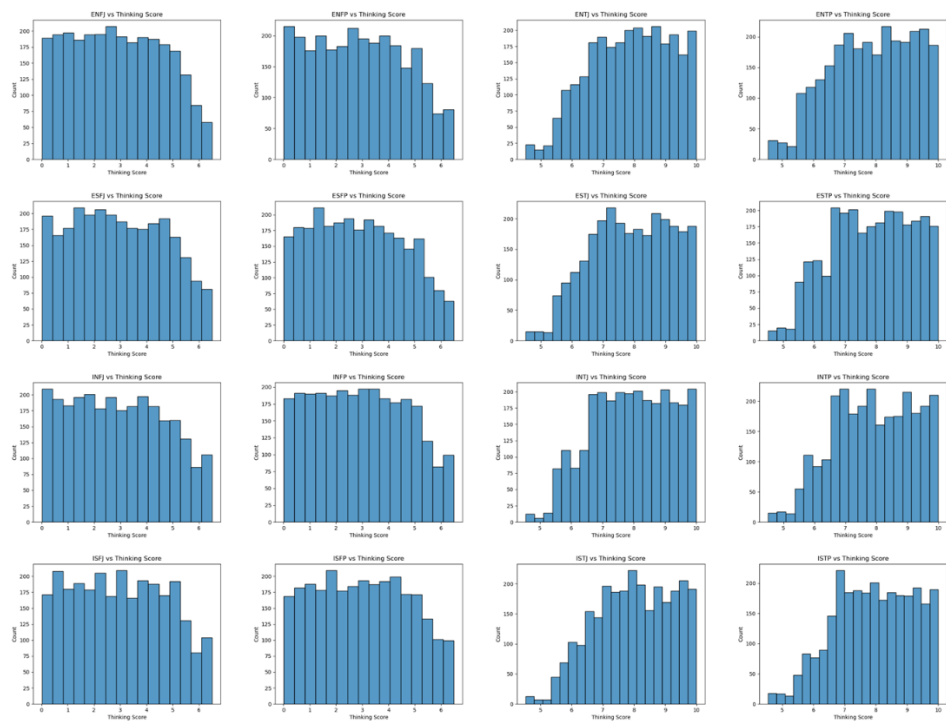


図 3-11 MBTI 別思考スコア

図 3-11 から T(思考型)はある地点からヒストグラムが右肩上がるがどこから上がるかが MBTI によって異なるのに対して F(感覚型)はある地点まで一様分布のような形をしてそこから右肩下がりヒストグラムになっている。

3.5 カーネル密度関数による MBTI 別スコアの可視化

カーネル密度関数(式 3.1)[11]で内向性・感覚性・思考型・判断型を可視化したグラフを図 3-12 に示す。

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \cdots (3.1)$$

$f(x)$:推定された密度関数

K :ガウシアンカーネル(正規分布)

h :バンド幅(スムージングパラメータ)

x_i :観測値

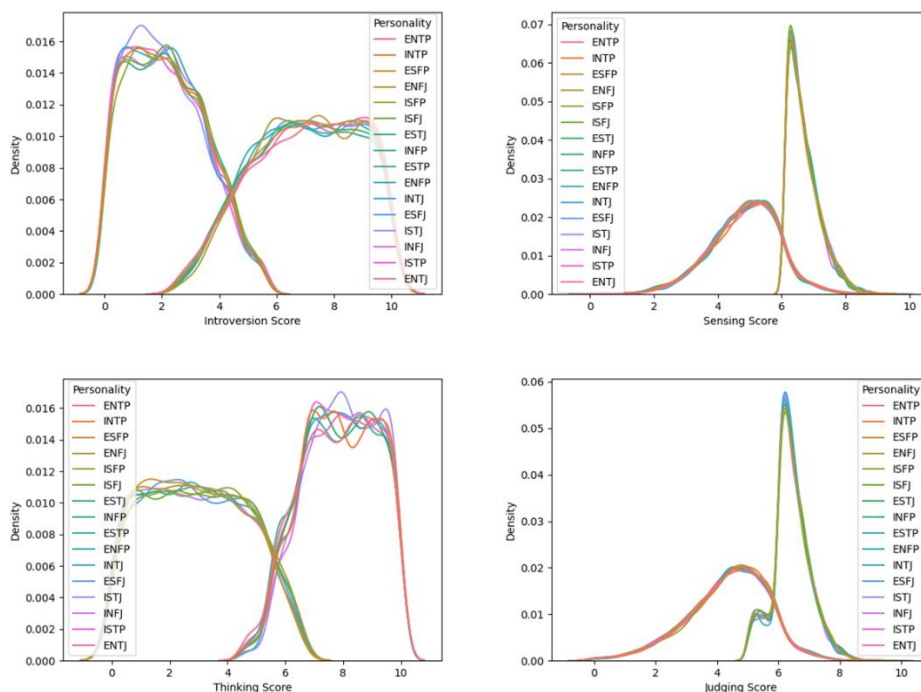


図 3-12 MBTI 別に内向性・感覚性・思考型・判断型のスコア分布をカーネル密度推定(KDE)で可視化

各ペアの分布が重なっている範囲に注目し、分類困難な性格ペアの視覚的傾向を示している。特にこの KDE プロットでは内向スコアと思考スコアおよび感覚スコアが分布の重なりから分類困難性が確認できる。

内向スコアおよび感覚スコアならびに判断スコアが対の要素大きく分布が重なっていることが分かる。内向スコアと思考スコアも同じような重なり方をしているが内向スコアは約 2 から 6(差分 4)の範囲で思考スコアは約 4 から 7(差分 3)の範囲で思考スコアは分布の重なりが内向スコアより小さいことが分かる。

4 クラスタについて検証

よりデータを深く分析するために生データ標準化(式 4.1)して主成分分析[12]を用いて次元圧縮しデータを 2 次元に可視化した。この時に第一主成分と第二主成分の因子負荷量(式 4.2)を分布とともにプロットする biplot の結果を図 4-1 に示す。また因子負荷量[11]の計算に必要な主成分分析における寄与率(r_i)は(式 4.3)で記す。

$$z = \frac{x_i - \bar{x}}{\sigma} \cdots (4.1)$$

z : 変換後の値

x : 実測値

σ : 生データの標準偏差

$$l'_{ij} = \sqrt{r_i} e_{ji} \cdots (4.2)$$

$$r_i = \frac{\lambda_i}{\sum_{k=1}^p \lambda_k} \cdots (4.3)$$

λ_i : i 番目の固有値

e_{ji} : 第 i 主成分に対応する第 j 変数の固有ベクトル成分

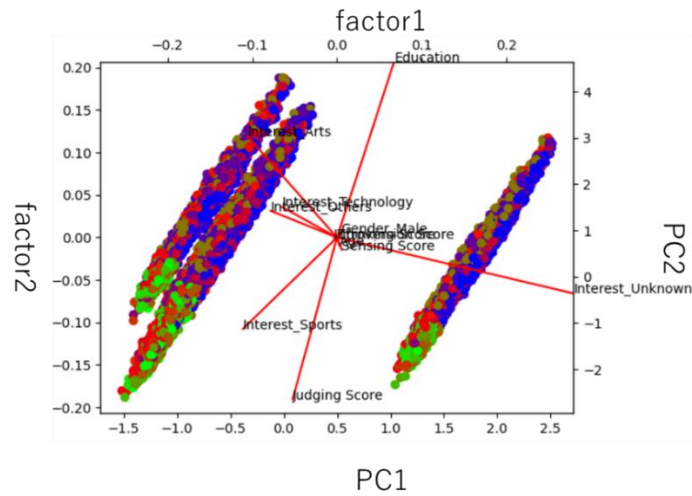


図 4-1 データセットの主成分分析と因子負荷量

また図 4-1 より 3 つのクラスタが確認できた。しかし MBTI についてはクラスタによって明確な傾向性(MBTI の値は色)が確認できないことが分かった。

biplot における因子負荷量の出力結果から横軸(第一主成分)は値が高いと Interest_Unknown が True(ダミー変数のため)になり Interest_Arts が False になる。また、因子負荷量から縦軸(第二主成分)は値が高いと教育の値が 1 になり、判断スコアが下がり Interest_Sports が False になる。

この biplot から興味(不明)で分離可能であることが分かる。

4.1 第三主成分までの因子負荷量

図 4-2 のクラスタリングは第三主成分まで使用している。そこで第三主成分までを可視化したグラフを図 4-3 に示す。

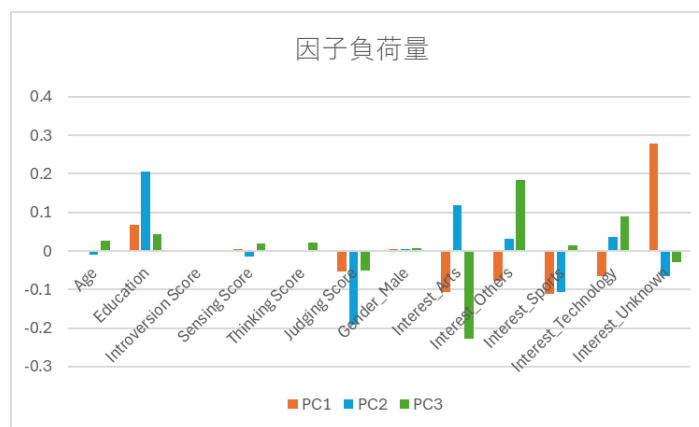


図 4-2 第三主成分までの因子負荷量

ここで、目立った第三主成分の因子負荷量については興味(芸術)が負に高く興味(他)と興味(科学技術)が正に高いことが分かる。

つまり第三主成分では値が正に大きいほど興味(他)・興味(科学技術)が **True** になり興味(芸術)が **False** になることが分かる。

5 考察

5.1 MBTI データセットについて

図 3-4 から図 3-11 において、MBTI では 16 性格ラベルの基準が明確にあり、それがデータセットに現れたことが表 3-2 における分類寄与率に現れたことが考えられる。

E(外向的)の性別は S(感覚型)か N(直感的)かで男女比が逆転しており TJ(論理的思考)の場合、他の S か N と比率が逆転している。また、N の時は男性が多く、S の時は女性が多い。また、I(内向的)の性別は INFJ と ISFJ で同じ性別の多さは同じ程度の比率だが E と比べて男女比が大きく異なる。同様に、INFP と ISFP は性別が逆転しているが E と比べて男女比が大きく異なる。

I と E の次に S と N が間違えている傾向にあることについて分類寄与率の各スコアが 4 つのスコアの中で最も低く一軸だけでの判断が難しかった事や図 3-12 からカーネル密度関数の分布が大きく重なっていることが要因として考えられる。

図 3-12 のカーネル密度関数から MBTI の対になる要素でも分布が一部被っているところがあり、これが精度を落とした要因と考えられる。

また、図 3-12 から判断スコアの KDE プロットでは大きく分布重なっている部分があるが、図 3-5 から教育について着目すると P が大学院まで行き J が大学院まで行かない傾向があることがわかり、J と P の精度を向上させた要因だと考えられる。

5.1.1 社会属性の必要性

本データセットは診断スコアと社会属性から成り立っている。そこで社会属性をなくして 500 回クロスバリデーションを行った。

その結果得られたようやく統計量を表 5-1 に記す。

表 5-1 CV 結果における診断スコアのみと社会属性ありの比較

	cross_validation_4	cross_validation_8
count	500.000000	500.000000
mean	0.739591	0.895525
std	0.047533	0.032741
min	0.602273	0.806818
25%	0.704545	0.873563
50%	0.738636	0.896552
75%	0.770115	0.919540
max	0.863636	0.988506

また、診断スコアのみ(4 変数)と社会属性を含むデータセット(8 変数)の KDE プロットと正規分布を図 5-1 に示す。どちらも Shapiro-Wilk 検定で $p < 0.05$ 未満で両方のデータのサンプル数は 500。

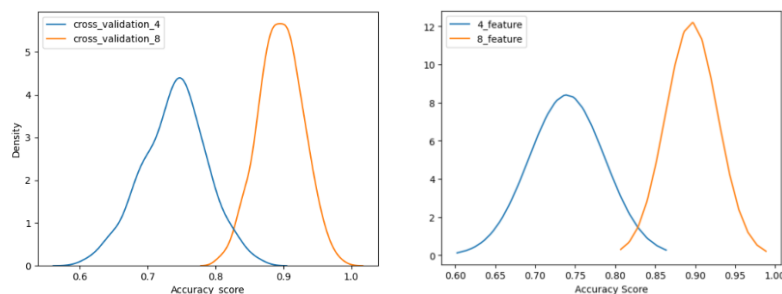


図 5-1 正解率の KDE プロットと正規分布

Shapiro-Wilk 検定では $p < 0.05$ だったが、KDE プロットから正規分布と比べて大きな歪みや偏りはないことが分かる。

ここでウェルチの t 検定を行った際の p 値は 2.49×10^{-316} で有意に診断スコアのみの場合と社会属性があるデータと比較して低かった。

また、誤分類について混合行列で作成したヒートマップを図 5-2 に示す。

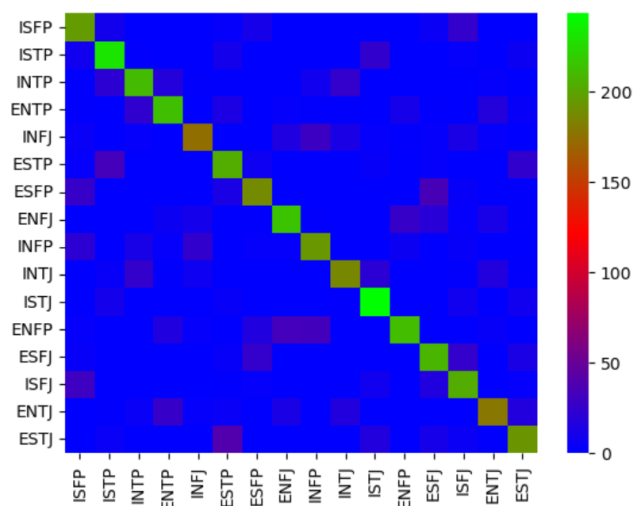


図 5-2 診断スコアのみの場合における混合行列のヒートマップ

図 3-3 および図 5-2 から社会属性の変数が無いと精度だけでなく誤分類傾向も把握しにくくなることが考えられる。

5.2 その他のクラスタ

第一主成分から第十二主成分までの分布を図 5-3 に示す。

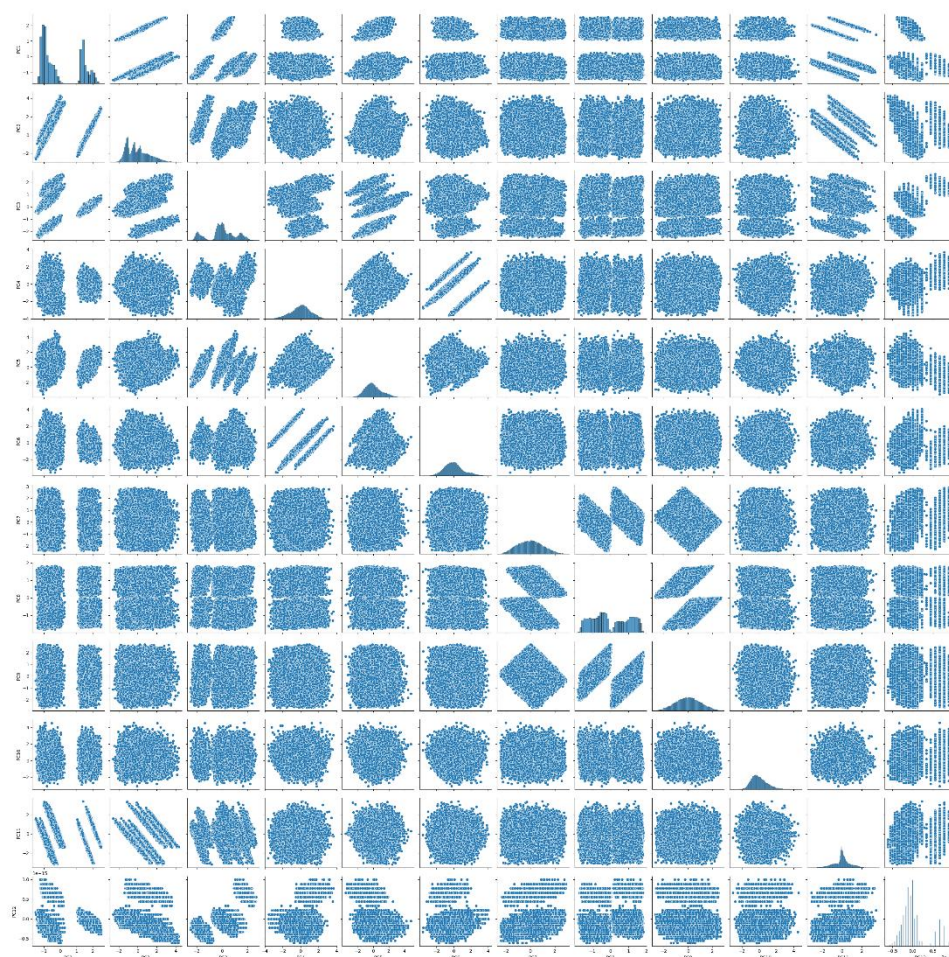


図 5-3 主成分分析後のペアプロット

第四主成分から第十二主成分までの因子負荷量を可視化したグラフを図 5-4 から図 5-6 に示す

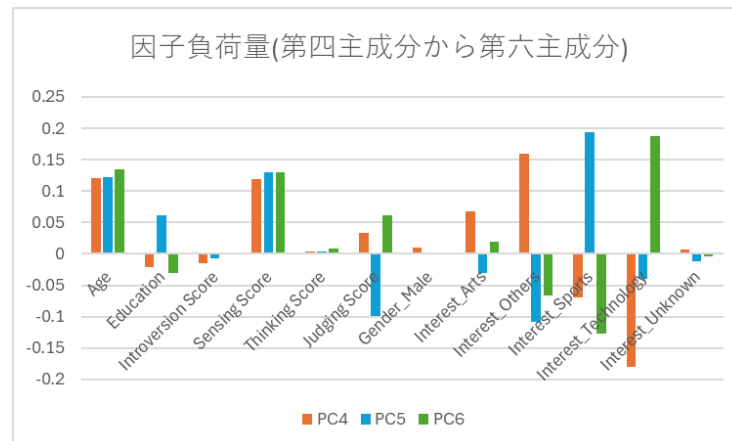


図 5-4 第四から第六主成分の因子負荷量

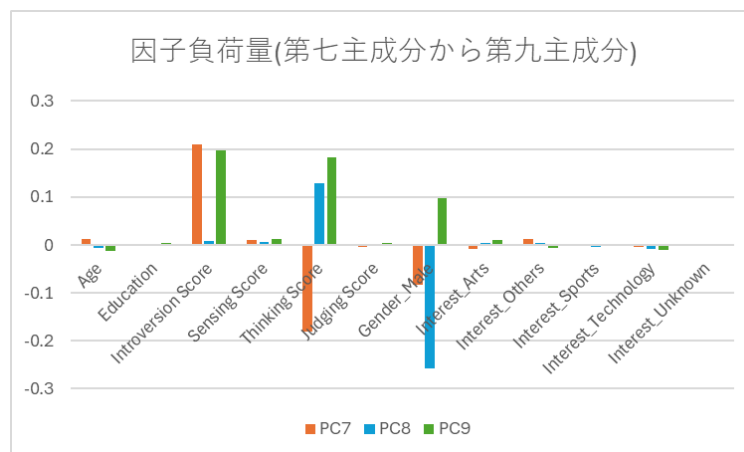


図 5-5 第七から第九主成分の因子負荷量

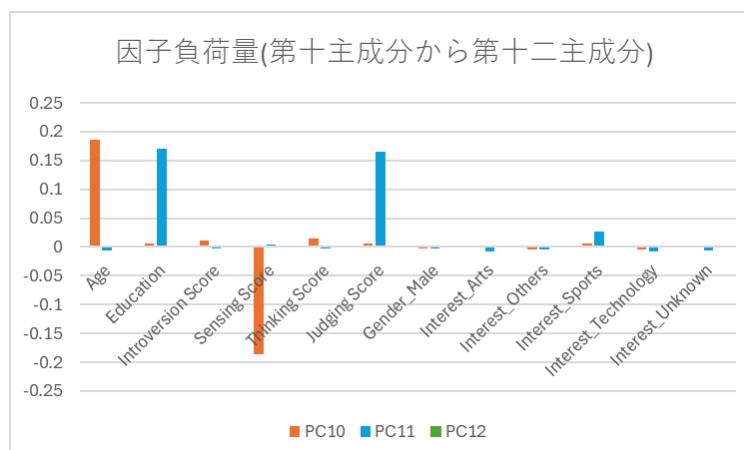


図 5-6 第十から第十二主成分の因子負荷量

図 5-3 における第七主成分と第八主成分でひし形のクラスタ(図 5-7)が確認できた。

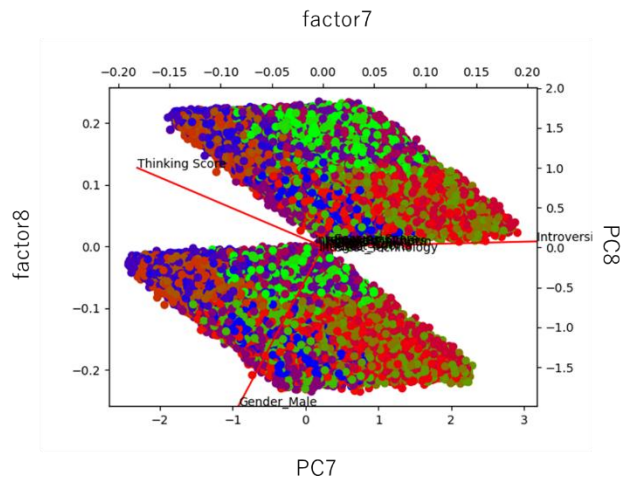


図 5-7 第七主成分と第八主成分の biplot

図 5-7 の散布図では縦方向のクラスタ分離が確認され、これに性別変数が関与していると考えられる。このクラスタは性別で分類可能クラスタで、第七主成分は値が大きいと内向スコアの値が高くなり思考スコアの値が低くなると考えられる。

その他に見られたクラスタとして長方形のクラスタ(図 5-8)が確認できた。

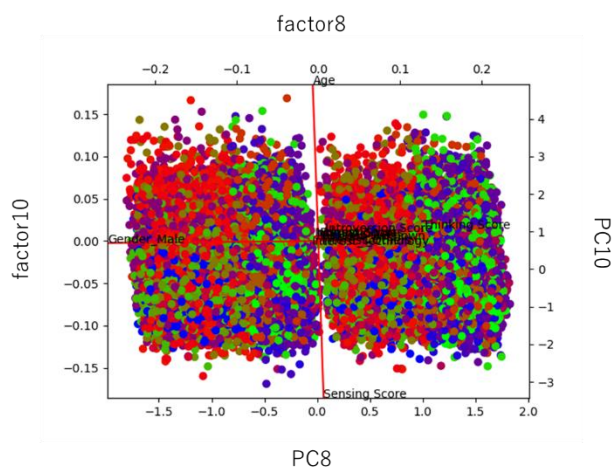


図 5-8 第八主成分と第十主成分の biplot

図 5-8 からクラスタを横方向のクラスタが分離が確認され、これに思考スコアと性別が関与していると考えられる。第十主成分からクラスタは値が高いと年齢が高く感覚スコアが低いと考えられる。

図 4-1 および図 5-7 並びに図 5-8 から共通して言えることとしてバイナリ変数がクラスタを生成していると考えられる。

他に図 5-3 におけるペアプロットの対角行列からヒストグラムが左右でまたはクラスタの要素数ごとに分かれていることが要因として挙げられる。

6 まとめ

MBTI の 93 個の質問が完璧だと仮定したとき、今回の 8 個の特徴量では代表値の中で最頻値を採用するとしたときに 10%程度精度が落ちてしまうことが分かった。

社会属性の変数である性別・興味・教育・年齢は精度を上げる要因となっており、診断スコアのみで MBTI を分類するのは限界があることが分かった。そのため、適切な社会属性変数を追加することでより精度が上がると言える。

主成分分析後のペアプロットから他にもクラスタが見られ特に第七主成分と第八主成分の散布図のクラスタ(図 5-7)は非常に規則性のある図形として現れかつ解釈可能性が高いクラスタだった。

4 変数と 8 変数(社会属性あり)でのクロスバリデーション結果で t 検定を行い極端に低い p 値になった要因としてはデータ数が 500 件あったため効果よりもサンプル数が多かったためと考えられる(Cohen's $d= 3.824$)。

本稿の目的は構造の可視化と相関傾向の把握であり、心理学的ラベリングは今後の課題とする。

7 補足

7.1 他のモデルの検討

比較手法としてランダムフォレスト[14]による MBTI 分類も実施した。その結果、平均精度は LightGBM (GBDT) と同等水準であったが、学習および予測に要する計算時間が非常に長く、PC への処理負荷も高かった。

そのため、本研究では実験環境の制約を考慮し、クロスバリデーションの繰り返し回数を LightGBM の 500 回に対し、ランダムフォレストでは 50 回に限定した。

精度の傾向比較においては、50 回の繰り返しで十分な評価が可能と判断している。ランダムフォレストによる精度評価の結果を表 7-1 に示す。

表 7-1 ランダムフォレストを用いた MBTI 分類精度(CV=50)

cross_validationn_RFC	
count	50.000000
mean	0.900809
std	0.011279
min	0.880000
25%	0.892479
50%	0.900571
75%	0.909714
max	0.930286

表 3-1 と比較して LightGBM より標準偏差が低く精度の分布が平均値周辺に集中している事が確認できた。

7.2 MBTI の妥当性に関する批判について

実際に MBTI は心理学的妥当性の面で一部から批判されており[15]、再現性や信頼性に課題があるという指摘がある。

しかし本研究では MBTI そのものの支持ではなく既存の MBTI スコアデータに基づき 93 個の設問を経由せず逆から診断可能かという工学的視点からの検証を行っている。

また、研究背景でも記載したが MBTI は教育や人事などにおいて幅広く実務的に活用されており、設問の簡略化に本研究の意義があると考ええる。

7.3 社会属性の使用とプライバシーの懸念

本研究では社会属性の使用の有無でクロスバリデーションの平均値と中央値が約 16%の差が確認され社会属性を使用することで分類精度の有意な上昇することが示された。

今後はプライバシー配慮への観点から、社会属性代替となる変数選択や匿名化技術の活用などについても検討する。

7.4 クラスタの心理的意味が曖昧である点

本研究では主成分分析による biplot から潜在的クラスタを可視化した。しかしクラスタの心理的あるいは社会的意味づけが完全に明確ではなく、biplot から読み取れる情報から分離可能性とクラスタの位置による意味の傾向を考察している。

しかし 4 章および 5.2 項で年齢以外の社会属性(バイナリ変数・ダミー変数)が分類に影響していることが視覚的に確認できた。

そのため MBTI 分類以外の社会的傾向の分析にも応用できると考える。今後はそのほかの次元圧縮アルゴリズムによる構造分析も検討する。

参考文献

- [1] Predict People Personality Types | Kaggle(2025 年 6 月 10 日アクセス)
<https://www.kaggle.com/datasets/stealthtechnologies/predict-people-personality-types?select=data.csv>
- [2] Google Scholar, “MBTI”, 検索結果 (2025 年 7 月確認) . https://scholar.google.com/https://scholar.google.com/scholar?hl=ja&as_sdt=0%2C5&q=MBTI&btnG=
- [3] CiNii, “MBTI”, 検索結果(2025 年 7 月確認) . <https://cir.nii.ac.jp/https://cir.nii.ac.jp/all?q=MBTI>
- [4] PubMed, “MBTI”, 検索結果(2025 年 7 月確認) . <https://pubmed.ncbi.nlm.nih.gov/https://pubmed.ncbi.nlm.nih.gov/?term=MBTI>
- [5] JSTOR, “MBTI”, 検索結果(2025 年 7 月確認) . <https://www.jstor.org/https://www.jstor.org/action/doBasicSearch?Query=MBTI&so=rel>
- [6] Harrington, Rick, and Donald A. Loffredo. "MBTI personality type and other factors that relate to preference for online versus face-to-face instruction." *The Internet and Higher Education* 13.1-2 (2010): 89-95.
- [7] Coe, Charles K. "The MBTI: Potential uses and misuses in personnel administration." *Public Personnel Management* 21.4 (1992): 511-522.
- [8] Harrington, Rick, and Donald A. Loffredo. "MBTI personality type and other factors that relate to preference for online versus face-to-face instruction." *The Internet and Higher Education* 13.1-2 (2010): 89-95.
- [9] Harvey, Robert J., William D. Murry, and Steven E. Markham. "Evaluation of three short-form versions of the Myers-Briggs Type Indicator." *Journal of Personality Assessment* 63.1 (1994): 181-184.
- [10] Ke, Guolin, et al. "Lightgbm: A highly efficient gradient boosting decision tree." *Advances in neural information processing systems* 30 (2017).
- [11] データ解析 第十回「ノンパラメトリック密度推定法」 | 山西健司(東京大学)
<https://ibis.t.u-tokyo.ac.jp/suzuki/lecture/2015/dataanalysis/L9.pdf>
- [12] Principal Component Analysis | SpringerLink
<https://link.springer.com/book/10.1007/b98835>
- [13] 主成分分析 | 統計科学研究所
https://statistics.co.jp/reference/software_R/statR_9_principal.pdf
- [14] Breiman, Leo. "Random forests." *Machine learning* 45 (2001): 5-32.
- [15] 若者層に大流行の「MBTI」に重大な誤解? 「16Personalities」の過信は人生の選択肢を狭める (東洋経済 education×ICT) - Yahoo!ニュース
<https://news.yahoo.co.jp/articles/13cc1ee19ceca2035a8671ed888d6d54c3ad3148>