Vol.2021-DBS-173 No.4
Vol.2021-IFAT-144 No.4
2021/9/16

分類可能性予測における学習データに求める性質の検討

早川 雄登[1,a)]　新美 礼彦[2,b)]

概要：近年，大規模データの処理方法の一つとしてデータマイニングが様々な分野で応用されている．データマイニングを効果的に運用するためには，分析対象データから有用な知識が得られるのかどうかを確かめる必要があり，適切な判断のためにデータマイニングに関する専門知識が必要である．しかし，データマイニングが一般的なタスクとなった弊害として，必ずしも分析を行う主体がその専門知識を持っていないという問題が生じている．そこで我々はデータマイニングの中でも分類タスクに着目して，分析対象データから分類タスクによって有用な知識が得られる期待度として分類可能性を提案し，専門知識に依らずに予測する方法を研究している．その試みの一つとして，分類可能性予測器の事前構築を行う方法について検討しているが，その学習データに求められる性質は不明である．本稿では，実際に分類タスクで用いられるデータを用いた分類可能性予測を行い，学習データに求められる性質について議論する．

Discussion on Training Data in Classificatability Prediction

1. はじめに

近年，様々な領域においてデータマイニングに注目が集まっている．特に社会の IoT 推進により大規模データが生成されているのに対し，効率的な処理方法としてデータマイニングを行うことにより蓄積されたデータから新たな知見を得られることが期待されている．

しかし，データマイニングを行いたい主体が必ずしもデータマイニングに関する知識を持っているとは限らない．これはデータマイニングが一般的な作業となったことの弊害であり，データマイニングの需要に対しデータサイエンティストの供給が追いついていないことに起因している．それにより，データマイニングの知識の欠如からデータの価値判断が行えずにデータマイニングへのコスト投入が求められることが考えられる．

また，どのようなデータであっても，データマイニングにより必ずしも有用な知識が得られることは保証されない．データマイニングは得られたデータから様々な方法を用いて法則性を発見し，それを解釈することで知見を得るタスクである．その際，対象としているデータにある法則性が成立していなければ発見することができない．

データマイニングを行うために生じる追加のコストは，データから知見を得るための投資であると考えられる．しかし，上述の通りその投資を行う時点ではデータからコストに見合った知見を得られるかは定かではない．分析の対象としているデータから知見が得られる期待度が高いほど投資を行いやすくなると考えられるため，そのデータを分析した際にどの程度パタンが発見できるかを簡易的に判断するシステムは有用であると考えられる．

本研究では，データマイニングの中でも分類問題に焦点を当てて，分類問題における対象データの期待度を「分類可能性」と定義し，それを予測するシステムを構築することを目的としている．

本稿では分類可能性予測システムにおいて，分類可能性予測器の学習データに求める性質について議論する．分類可能性は 2 値で定義しており，分類可能性予測器はデータセットのメタ特徴からそのデータセットの分類可能性を分類予測する．分類可能性予測器を構築する際に学習データとしてあるデータセットのメタ特徴と分類可能性の組が必要となるが，この学習データに求められる性質は明らかではない．そのため，実データを用いた分類可能性予測の結果を元にそれらのデータの性質について議論する．

[1] 公立はこだて未来大学大学院 システム情報科学研究科
Future University Hakodate,
Graduate School of Systems Information Science
[2] 公立はこだて未来大学 システム情報科学部
Future University Hakodate,
Faculty of Systems Information Science
[a)] g2120036@fun.ac.jp
[b)] niimi@fun.ac.jp

2. 人工データを用いた分類可能性予測

早川ら [1] は，あるデータセットに対し分類タスクを行なった際に，そのデータセットからパタンが発見できる期待度を図る指標として分類可能性を定義した．さらにデータセットの性質から分類問題としてそれを予測する方法を提案し，人工データを用いた実験によりその有効性を示した．

しかし，早川らの論文では実データを用いた試行はしておらず，実用可能性が示されていない．また，人工データ生成は乱数によるものであるが，いくつかの乱数が複合しているため人工データに対する議論が困難である．

そのため，ここでは 2 つの実験を行う．1 つ目は再現実験として後述する 3 種類の方法を用いてそれぞれ人工データを生成し，それらの人工データのみを用いたものである．2 つ目は早川らの論文で示唆されていた人工データによる学習により構築されたモデルで実データの分類可能性の予測するものである．

2.1 人工データ生成方法

3 種類の人工データはそれぞれ以下のような性質を持つ．

- データ群 1
  数値 ($\mathcal{N}(0,1)$ から生成) のみ
- データ群 2
  バイナリ値 (等確率の二項分布から生成) のみ
- データ群 3
  数値とバイナリ値の組み合わせ (比率は 1 : 9 から 9 : 1)

ここで，数値は Stevens の尺度分類 [2] のうち間隔尺度を持つもの，バイナリ値は名義尺度あるいは順序尺度を持ち間隔尺度を持たないものを指している．

また，カテゴリカルデータについてバイナリ値のみを用いるが，これは多値カテゴリに対する一般的なアプローチとしてダミーコーディング [3] や誤差修正コーディング (Error-Correcting Code)[4] 等の数値化アプローチが取られた際に，処理後の個々の属性はバイナリとなるためである．ただし，特にダミーコーディングが行われた場合，複数カテゴリが等頻度となる分布を持っていたとしても処理後の属性は等頻度とならないことに注意が必要である．

これら 3 種類それぞれに対し属性数および事例数，データ群 3 においては比率を変化させたいくつかの人工データセットを作成する．各データセットでは各係数を正規分布により決定した線形結合によって得られた値を閾値 0 で 2 値化したものをクラス属性とする．その後，事例と属性を削減することにより分類性能を変化させる．

2.2 予測に用いるメタ特徴

データの性質として，データセットのメタ特徴を抽出する．そこで用いるメタ特徴は早川らの論文に準ずるが，その中に数値属性を含まないデータセットに対してはそのままでは抽出できないものが含まれる．そのため，早川らの論文で提示された 18 種類のメタ特徴から以下の 9 種類のメタ特徴を用いる．

(1) インスタンス数 (nr_instances)
(2) データセットの次元性 (dimensionality)
(3) 目的属性の情報エントロピー (class_ent)
(4) 最大エントロピーで正規化された各属性の情報エントロピーの平均値 (norm_ent)
(5) 目的属性に対する相互情報量の平均値 (MI_max)
(6) 目的属性に対する相互情報量の最大値 (MI_avg)
(7) 等価特徴数 (EqFeats)
(8) 雑音信号比 (NSR)
(9) Fisher 判別比の最大値 (FDRatio_max)

数値属性を含まないデータセットに対して抽出ができないメタ特徴を不採用にした理由は，カテゴリカルデータは間隔尺度を含まないため，数値化を行ないメタ特徴を抽出するとデータの性質とは言い難いものとなるためである．

2.3 人工データのみを用いた試行

前述の方法により生成した人工データ群について，それぞれ分類可能性とメタ特徴を抽出しメタ特徴・分類可能性データセットを作成する．そして，それを用いて Random Forest により分類可能性予測器を構築，10-Fold 交差検証により評価を行う．実験は scikit-learn[5] を用いて行い，Random Forest のハイパーパラメタは scikit-learn の標準パラメタを用いる．

この実験では，まず学習データの事例数全てを用いて構築と評価を行ったが，いずれのデータ群においても高い予測性能が得られなかった．これは早川らの行った実験でも述べられていたが，この問題の原因は生成された人工データの多くがここで用いている分類可能性の定義においては分類に適しておらず，学習データの不均一性が生じていることによるものであると考えられる．

そのため，メタ特徴・分類可能性データセットに対して Random Under Sampling を行なった後に分類器構築・評価をした結果を表 1 から表 3 に示す．

ここで，評価指標として “F1(正)” と “F1(負)” を提示しているが，これらはそれぞれ正例に対する F1 値と負例に対する F1 値である．正例に対する F1 値と負例に対する F1 値は，Confusion Matrix における TP，TN，FP，FN を用いて下式によって求められる．

$$F1(\text{正}) = \frac{2TP}{2TP + FP + FN}, F1(\text{負}) = \frac{2TN}{2TN + FP + FN}$$

結果的に全てのデータ群において良好な分類が可能であった．特にデータ群 1 とデータ群 2 では Accuracy，正負の F1 値について 8 割以上の分類性能が得られた．

表 1 データ群 1 の各閾値における予測性能 (対処後)

閾値	正例	負例	正例率	Accuracy	F1(正)	F1(負)
0.65	1431	1431	50.0%	0.867	0.867	0.867
0.70	515	515	50.0%	0.932	0.932	0.932
0.75	181	181	50.0%	0.961	0.961	0.961
0.80	61	61	50.0%	0.975	0.976	0.975
0.85	11	11	50.0%	0.909	0.917	0.900

表 2 データ群 2 の各閾値における予測性能 (対処後)

閾値	正例	負例	正例率	Accuracy	F1(正)	F1(負)
0.65	1392	1392	50.0%	0.804	0.807	0.800
0.70	761	761	50.0%	0.798	0.803	0.792
0.75	389	389	50.0%	0.802	0.803	0.802
0.80	185	185	50.0%	0.838	0.835	0.840
0.85	69	69	50.0%	0.819	0.823	0.815

表 3 データ群 3 の各閾値における予測性能 (対処後)

閾値	正例	負例	正例率	Accuracy	F1(正)	F1(負)
0.65	1798	1798	50.0%	0.627	0.635	0.619
0.70	911	911	50.0%	0.659	0.662	0.657
0.75	385	385	50.0%	0.681	0.685	0.675
0.80	151	151	50.0%	0.666	0.660	0.671
0.85	50	50	50.0%	0.670	0.660	0.680

また，データ群 3 においてデータ群 1，2 と比べて高い分類性能が得られなかったことについては，Random Under Sampling によりデータバリエーションが減少したことに起因していると考えられる．顕著にその影響を受けた理由は，データ群 3 は単に数値やバイナリデータを用いたのではなくそれらを複合したことであると考えられる．そのため，より複雑なデータを用いる際には学習データのバリエーション確保が必要であると考えられる．

ここで，Random Under Sampling を行った後の各データセットのメタ特徴の散布図を図 1 に示す．この図では，各属性について各データ群における分布を示しており，さらに各データ群における中央値を水平線で示している．

これらのメタ特徴のうち，データ群 1 における class_ent と norm_ent，FDRatio_max，データ群 2 における norm_ent，EqFeats，NSR がそれぞれほぼ同一の値となった．この理由として，データ群 1 は各属性を標準正規分布から生成した上で標準正規分布より生成された係数を用いた線形結合を行うことでクラス属性を作成していることでクラス属性に正規分布が保持されるため，データ群 2 は各属性を 2 項分布から生成した結果 norm_ent がごくわずかとなるためであると考えられる．

2.4 実データへの適用

前述の各人工データ群から抽出したメタ特徴・分類可能性データセットを用いて分類可能性予測器を構築する．そして，実データから抽出したメタ特徴・分類可能性データセットを用いて分類可能性予測器で予測する．また 2.3 節で述べた結果を基に，学習データに対して Random Under Sampling を用いる．

表 4 実データに対するデータ群 1 の各閾値における予測性能

閾値	正例	負例	正例率	Accuracy	F1(正)	F1(負)
0.65	158	96	62.2%	0.583	0.731	0.070
0.70	147	107	57.9%	0.622	0.564	0.667
0.75	137	117	53.9%	0.531	0.270	0.655
0.80	124	130	48.8%	0.630	0.484	0.712
0.85	108	146	42.5%	0.701	0.563	0.772

表 5 実データに対するデータ群 2 の各閾値における予測性能

閾値	正例	負例	正例率	Accuracy	F1(正)	F1(負)
0.65	158	96	62.2%	0.437	0.565	0.201
0.70	147	107	57.9%	0.398	0.526	0.173
0.75	137	117	53.9%	0.350	0.483	0.127
0.80	124	130	48.8%	0.437	0.563	0.210
0.85	108	146	42.5%	0.421	0.491	0.329

表 6 実データに対するデータ群 3 の各閾値における予測性能

閾値	正例	負例	正例率	Accuracy	F1(正)	F1(負)
0.65	158	96	62.2%	0.421	0.502	0.310
0.70	147	107	57.9%	0.461	0.568	0.283
0.75	137	117	53.9%	0.445	0.515	0.350
0.80	124	130	48.8%	0.457	0.552	0.310
0.85	108	146	42.5%	0.484	0.604	0.260

ここで，実データは UCI Machine Learning Repositories[6] で予測対象が nominal なデータセットから 71 種類を用いる[*1]．これらのデータセットには多クラス分類問題がいくつか含まれているため，メタ特徴および分類可能性を抽出する前に 1–vs–other 問題として分割する．

予測結果の性能評価を表 4 から表 6 に示す．

全てのデータ群において高い分類性能は得られなかった．その原因は大きく分けて 3 つ考えられる．

1 つ目は，生成した人工データが予測に相応しくないということである．この実験で使用した人工データは，2.1 節で述べた 3 つの性質をそれぞれ持つものである．しかしこれらは生成背景が実際に存在しているデータに比べて単純である恐れがあり，実データの予測に対しては性質が欠如していると考えられる．そのためこれが原因ならば，学習データに求める性質について検討することで解決できる可能性がある．

2 つ目は，分類可能性予測を行うためにメタ特徴が不足しているということである．この実験では，2.2 節で述べた 9 つのメタ特徴を使用した．これらのメタ特徴は，主に Hutter らの書籍 [8] において一般的なメタ学習タスクにおいて用いられるものであり，人工データによる実験において効果的であったため利用している．しかし，ここで対象としている実データに対しては有効性の確認ができていない．そのためこれが原因ならば，メタ特徴の選出を再度行う必要があると考えられる．

3 つ目は，この方法で分類可能性を予測することが困難

[*1] weka 用に GitHub で配布されているもの [7] より

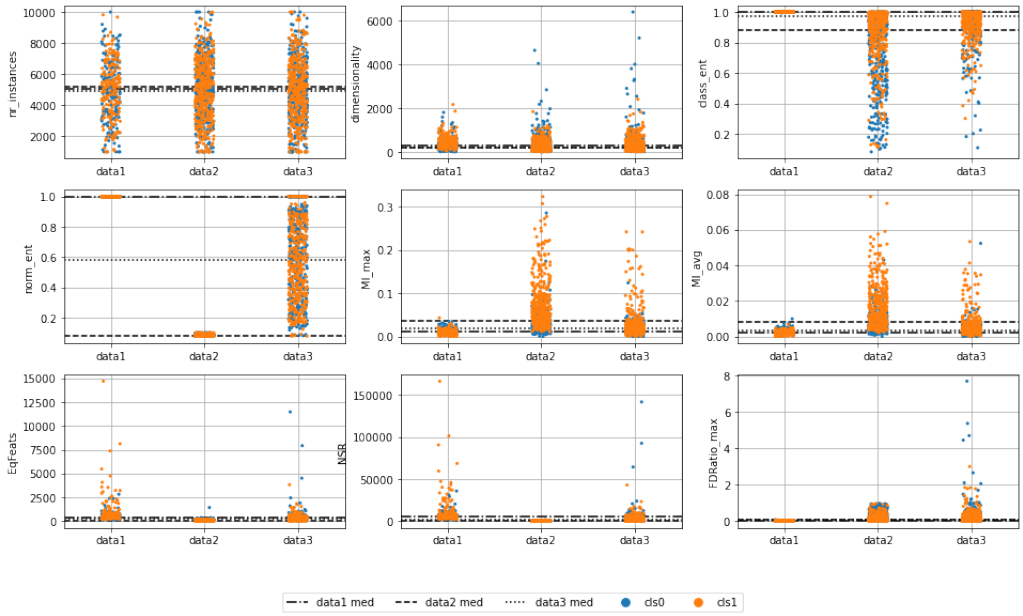

図 1　人工データ (対処後) のメタ特徴分布 ($\theta = 0.75$)

であるということである．この方法は早川らの論文および 2.3 節において人工データに対して有効であることは確かめられているが，ここで対象としている実データに対しては有効性の確認ができていない．そのためこれが原因ならば，この方法の限界に関する議論を行なった上で，その限界に沿ったデータにのみ適用すべきであると考えられる．

3. 半データ学習モデル

2.4 節において述べた原因のうち 1 つ目を確認するために “半データ学習モデル” を提案する．

“半データ学習モデル” は分析対象データセットを分割して分析・検証する方法であるが，交叉検証など分析対象データセットのサブセットをそのまま用いて学習器を構築する諸手法とは異なる．この方法はメタ特徴を用いた学習に特化したものであり，分析対象データを予め分けた上でメタ特徴抽出を行うことで，類似したデータセットのメタ特徴を得ることを目的としたものである．同じデータセットから抽出したメタ特徴を 2 度用いることとの違いは，分割後のデータセットのメタ特徴は必ずしも元のデータセット及びもう一方の分割後データセットとは一致しないと考えられるためである．しかし，全く異なるデータを用いることと比べると母集団が同一であると考えられるため類似したメタ特徴を得られることが期待される．

この方法を用いて，分類可能性予測を行うための学習データとして予測したいデータに近い性質を持ったデータのメタ特徴・分類可能性データセットを利用することにより予測が可能であるかということを確かめる．それが可能であれば，2.4 節で高い分類性能が得られなかった原因の 1 つは用いた学習データにあると考えられる．

まず，あるデータセットを分割後のデータセットのクラス分布が分割前のデータセットと同程度となるように二等分する．クラス分布が同程度となるように分割する目的は，分割後の両データセット間で分類可能性が同程度となることを狙うためである．また分割時の制約条件をクラス分布以外に置かないことで，同一母集団からサンプリングされたデータセットであるが必ずしもメタ特徴が一致しないことが期待されるが，クラスエントロピーなど一部のメタ特徴はこの制約条件に強く影響を受けることは留意すべきである．

次に，分割後の両データセットに対し，それぞれメタ特徴と分類可能性を抽出する．この手順は**??**，2.2 節で述べたものと同じである．

最後に，得られた一方のメタ特徴・分類可能性データセットで学習した分類可能性予測器によりもう一方の予測を行う．これにより，同一母集団から抽出されたデータセットのメタ特徴・分類可能性データセットで学習したモデルによる予測が可能であるかを確かめられる．

二等分したそれぞれのデータ $i(i = 1, 2)$ 群のメタ特徴セット X_i，データ i 群の各データにおける各分類評価指標の最大値を閾値 θ で分類可能性とした時 $y_{i;\theta}$，またその予測値 $\hat{y}_{i;\theta}$ と表記した時，モデル M_θ は X_1 と $y_{1;\theta}$ により学習し，M_θ に X_2 を入力して得られた $\hat{y}_{2;\theta}$ を $y_{2;\theta}$ と比較することで評価する方法を，半データ学習モデルとする．

4. 半データ学習モデルの性能と振る舞い

3 節で述べた半データ学習モデルを用いた実験の結果を表 7 から表 9 に示す．実験に用いたデータセットは 2.4 節で述べたものと同様であるが，データセットを分割する際

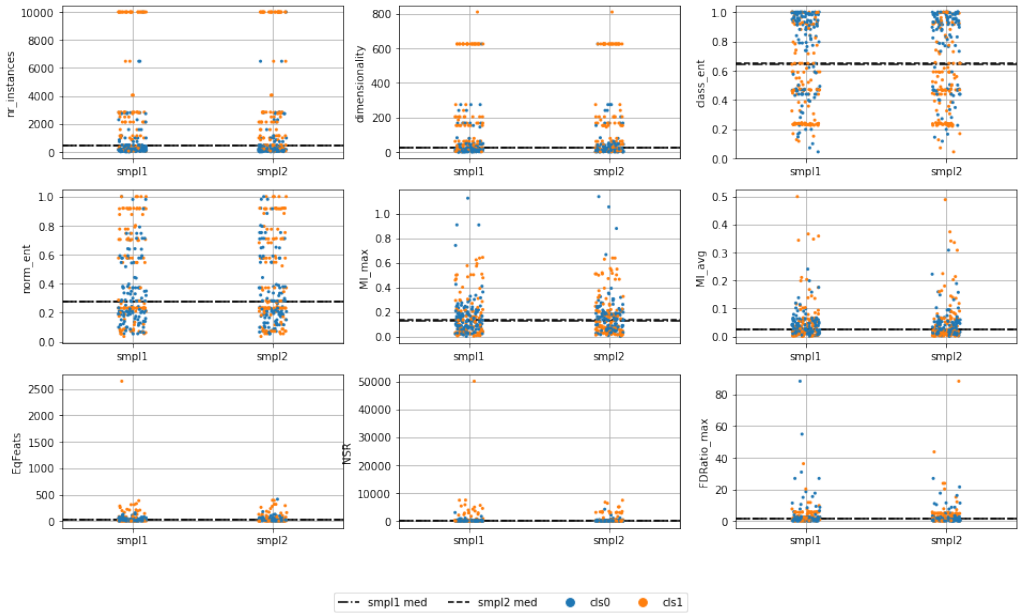

図 2 実験で用いた実データのメタ特徴分布 ($\theta = 0.75$)

表 7 データ群全数による結果

閾値	正例	負例	正例率	Accuracy	F1(正)	F1(負)
0.65	147	92	61.5%	0.812	0.834	0.783
0.70	136	103	56.9%	0.816	0.823	0.809
0.75	125	114	52.3%	0.824	0.822	0.826
0.80	110	129	46.0%	0.820	0.798	0.838
0.85	91	148	38.1%	0.833	0.759	0.872

表 8 一致群による結果

閾値	正例	負例	正例率	Accuracy	F1(正)	F1(負)
0.65	124	79	61.1%	0.916	0.929	0.897
0.70	110	91	54.7%	0.925	0.931	0.919
0.75	103	104	49.8%	0.937	0.937	0.938
0.80	88	115	43.3%	0.906	0.890	0.918
0.85	72	134	35.0%	0.927	0.892	0.945

にあるクラスに属するデータが 1 事例しかなかった場合は等頻度分割ができないため除外した．

ここでは，3 つの方針で実験を行なった．1 つ目はデータ群全てを用いたものであり，データ群全数とする．2 つ目はデータ群のうち $y_{1;\theta} = y_{2;\theta}$ であるものであり，一致群とする．3 つ目はデータ群のうち $y_{1;\theta} \neq y_{2;\theta}$ となるものであり，不一致群とする．また，2 つ目と 3 つ目では各閾値における一致を考えるため，正例と負例の事例数が変化する．

データ群 1 とデータ群 2 は同母集団から抽出した別データと考えられる．特に一致群では，一方のデータセットを用いて構築した分類器はもう一方のデータセットを用いて構築した分類器と同等以上の性能を示すと考えられる．それに対して，不一致群ではこれが保証されないと考えられる．

データ群全数 239 件に対し，一致群は約 200 件，不一致群は約 40 件となった．

表 9 不一致群による結果

閾値	正例	負例	正例率	Accuracy	F1(正)	F1(負)
0.65	23	13	63.9%	0.222	0.067	0.333
0.70	26	12	68.4%	0.158	0.059	0.238
0.75	22	10	68.8%	0.156	0.129	0.182
0.80	22	14	61.1%	0.167	0.211	0.118
0.85	19	14	57.6%	0.061	0.000	0.114

データ群全数では予測性能は概ね良好であり，一致群では更に向上した．一方不一致群ではほとんど予測が不可能となったが，これは前述の通りどう母集団から得られたデータセットであっても予測性能が一致しなかったものに関する分類可能性予測であるため，想定通りの結果である．

これら結果から，データのサンプリングに偏りが生じていなければ，同じ母集団から抽出されたデータを用いて学習したモデルにより予測は可能であると考えられる．従って，想定されるデータの性質を発見し，それに則したメタ特徴・分類可能性データセットを予め用意できれば，想定した範囲内での分類可能性予測は可能であると考えられる．

なお，このデータの性質はメタ特徴で表現可能なものであり，分類可能性が異なるデータとの判別が可能であるものである必要がある．

5. 学習データに求める性質の検討

図 2 は実験に用いた実データのメタ特徴の散布図である．図 1 と比較すると，2.3 節で言及した人工データ群 1 および人工データ群 2 に見られるメタ特徴の傾向は実データにおいて見られないことがわかる．このことから 2.4 節で人工データにより学習した分類可能性予測器による予測で高い性能が得られなかった原因の一つは学習データによるものと考えられる．

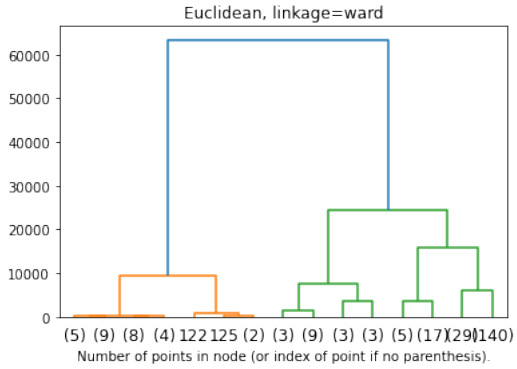

図 3 クラスタの樹形図 ($\theta = 0.75$)

表 10 各クラスタの正例負例分布 ($\theta = 0.75$)

クラスタ	P1	N1	P2	N2
1	9	10	8	11
2	7	0	7	0
3	9	2	9	2
4	16	13	19	10
5	39	90	45	84
6	25	1	25	1
7	2	2	2	2
8	1	0	1	0
9	2	0	2	0
10	3	8	7	4

次に，各データセットの傾向を示すクラスタが得られることを期待して，階層的クラスタリングを用いて 2.2 節のメタ特徴によりデータセット群のクラスタリングを行った．階層的クラスタリングの樹形図を図 3 に示す．その結果の一例としてクラスタ数が 10 のときの結果を表 10 に示す．

表は各クラスタにおける各データ群の分類可能性の分布を示している．表において P1 は分割データ群 1 で分類可能性が正となったものであり，N2 は分割データ群 2 で分類可能性が負となったものである．

クラスタリング手法として階層的クラスタリングを用いた理由は，クラスタ数の多少に応じて凝集度を調整できることである．しかし，その調整によって正例と負例の切り分けが困難であるクラスタの分解が行われていないことから，今回用いたメタ特徴およびその組み合わせではここで考えたいデータの性質を調べるために不足していることが考えられる．

これらのクラスタのうち，比較的綺麗な切り分けが行われたクラスタ 2，3，6，8，9 について考える．これらのうち 6 を除いた 4 クラスタはほぼ同様に分布しており，全データに対し正規化エントロピー (norm_ent) が高いという傾向を示した．これらの 4 クラスタは全て分類可能性が高いデータセットを多く含んでおり，高い norm_ent はその性質の一つであると考えられる．クラスタ 6 については他のデータセットに対し極めて高いインスタンス数 (nr_instances) と次元性 (dimensionality) を持つ．この原因は，2.4 節で述べた他クラス分類問題を 1–vs–other 問題として分割したことであり，クラスタ 6 に含まれるデータセットは全て同じデータセットより生成されているためであると考えられる．

次に，正例と負例が混在したクラスタ 4，5 について考える．これらのクラスタでは，正例と負例がほぼ同じ分布で混在している．そのため正例と負例をクラスタリングにより切り分けることは困難だと考えられる．しかし，用いるメタ特徴を変更することで正例・負例となるデータの性質をクラスタリングにより調べられる可能性がある．

6. おわりに

本研究では，「分類可能性」を予測するシステムを構築することを目的としている．その中で，本稿では分類可能性の予測方法と予測器の学習データに求める性質について議論し，分類可能性の予測可能性を示した．

先行研究に対し本稿で試行した人工データを用いた学習モデルによる実データの分類可能性予測は期待された結果が得られなかったが，実データの半データ学習モデルを用いた試行では高い予測性能を達成できたことから，人工データが学習データに求められる性質を満足できていないのではないかと考えられる．

半データ学習モデルを用いた試行の結果を基に，人工データに求められる性質を求める試みは，本稿で用いたメタ特徴では達成できなかった．

今後，分類可能性予測システム構築のために，分類可能性予測器の学習データを生成する方法について議論し，学習データ生成システムの構築を行うことを目標とする．

参考文献

[1] 早川雄登, 新美礼彦, “メタ特徴を用いた分類可能性予測”, DEIM Forum 2021 E11-4, pp. 1–6, 2021.
[2] S.S. Stevens, “On the Theory of Scales of Measurement.”, Science, Vol. 103, No. 2684, pp. 677–680, 1946.
[3] S. Garavaglia, A. Sharma, “A Smart Guide to Dummy Variables: Four Applications and a Macro”, Proceedings of the Northeast, Vol. 43, pp. 1–11, 1998.
[4] T.G. Dietterich, G. Bakiri, “Solving Multiclass Learning Problems via Error-Correcting Output Codes”, Journal of Artificial Intelligence Research, Vol. 2, pp. 263–286, 1994.
[5] “Scikit-Learn: Machine Learning in Python – Scikit-Learn 0.24.1 Documentation”, https://scikit-learn.org/stable/ (Accessed Aug. 9, 2021).
[6] D. Dua, C. Graff, “UCI Machine Learning Repository”, http://archive.ics.uci.edu/ml/ (accessed 2020-07-18), University of California, Irvine, School of Information and Computer Sciences, 2019.
[7] “Datasets – Weka Wiki”, https://waikato.github.io/weka-wiki/datasets/ (Accessed Aug. 9, 2021).
[8] F. Hutter, L. Kotthoff, J. Vanschoren, “Automated Machine Learning”, Springer, 2019.