

令和 6 年 5 月 27 日現在

機関番号：14401

研究種目：挑戦的研究（萌芽）

研究期間：2020～2023

課題番号：20K21815

研究課題名（和文）教師ラベル無しビッグデータからの高速高精度分類器学習手法の探求

研究課題名（英文）Study on fast and accurate classifier learning method from unlabeled big data

研究代表者

鷲尾 隆（Washio, Takashi）

大阪大学・産業科学研究所・教授

研究者番号：00192815

交付決定額（研究期間全体）：（直接経費） 4,900,000円

研究成果の概要（和文）：AI技術の普及に伴い、データ収集の制約やコストから教師ラベルなしビッグデータからの分類器学習ニーズが高まっている。これに対し、正負例割合の異なる2つのラベル無し事例集合から分類器を学習するUUC手法が提案されているが、既存手法はビッグデータに対して膨大な学習計算量を要し、また分類に偏り誤差を生じる問題があった。

そこで、本研究ではより汎用性を有し、低計算量で偏り誤差の無いUUC手法を提案した。そして、実データを含む様々なデータに適用し、教師ラベルあり学習と同等以上の精度で、教師ラベルなし学習が可能であることを検証した。これにより、既存手法の適用範囲限界を遥かに超えるUUC手法を確立した。

研究成果の学術的意義や社会的意義

IoT社会の深化とAI技術の普及に伴い、ビッグデータからの分類器学習ニーズが増しているが、多くの場合にデータ収集の制約やコストから教師ラベルが得られないことが問題となっている。これに対し近年、正負例割合の異なる2つのラベル無し事例集合から分類器を学習するUUC手法が提案されている。しかし、これらはカーネル法を用いており、訓練データ数Nについて $O(N^3)$ の学習計算量を要し、またNでも分類に偏り誤差を生じる場合がある。従って、複雑な事例分布を持つビッグデータに適用可能な高速高精度なUUC手法の開発が強く待たれていた。本研究成果は、この社会的要請に応えるものである。

研究成果の概要（英文）：With the widespread adoption of AI technology, there is an increasing demand for classifier learning from unlabeled big data due to constraints and costs associated with data collection. In response to this issue, the UUC method, which learns classifiers from two unlabeled datasets with different proportions of positive and negative examples, has been proposed. However, existing methods require vast computational resources for large-scale data and suffer from bias error in classification.

In this study, we propose a versatile UUC method which requires low computational cost only, and is free from bias error. We applied this method to the classification of various datasets, including real data, and verified that unsupervised learning without teaching labels is possible with almost the same accuracy as supervised learning. This establishes a UUC method that far exceeds the application range limitations of the existing UUC methods.

研究分野：人工知能

キーワード：弱教師有り学習 分類器学習 機械学習 UUC 教師ラベル無しデータ

1. 研究開始当初の背景

IoT 社会の深化と AI 技術の普及に伴い、ビッグデータからの分類器学習ニーズが増しているが、多くの場合にデータ収集の制約やコストから教師ラベルが得られないことが問題となっている。たとえば、地域住民の心機能健診データから個人の心不全発症リスクの高低を早期識別する分類器を得たくとも、個々人の心不全発症の長期追跡調査無しにはリスク高低の教師ラベル付けは困難である。また、微小生体の形状画像から種類識別する分類器を得たくとも、個体の各種類への精製が困難であれば正確な教師ラベル付けは難しい。ただし、心不全発症率が異なる地域や職業でデータを分ける、各種類の精製率を変えたデータを得る等、正負例割合が異なるラベル無しビッグデータを得ることは比較的容易な場合が多い。

従来、正負の教師ラベル付きとラベル無しの2つの事例集合を用いる半教師あり学習や、正ラベル付き事例集合 P とラベル無し事例集合 U から分類器 C を学習する PUC 手法が研究されている。これに対し近年、上記のように教師ラベルが皆無の場合、正負例割合の異なるラベル無し事例集合 U_1, U_2 を用いて分類器 C を学習する UUC 手法が提案されつつある [du Plessis13, Kaji19]。しかし、これらはカーネル法を用いており、訓練データ数 N について $O(N^3)$ の学習計算量を要し、また $N \rightarrow \infty$ でも分類に偏り誤差を生じる場合がある (不偏推定・一致推定が困難)。従って、ビッグデータに適用可能な高速性と偏りの無い高精度性を有する UUC 手法の開発が強く待たれる。

2. 研究の目的

提案者は、機械学習の分類問題に於ける分類境界は、分類対象事例が正例である確率と負例である確率の大小が反転する面であること、そして正負例割合の異なるラベル無し事例集合 U_1 と U_2 の確率密度の大小関係が反転する境界が U_1, U_2 それぞれの正負例割合に関わらず不変であり、かつ正例データと負例データの確率密度の大小関係反転面と一致することに鑑み、事例集合 U_1, U_2 から2つの確率密度の大小関係反転面を容易かつ正確に学習する手法を実現可能ではないかという着想に至った。本研究ではこの着想に基づき、従来の UUC 手法の問題点を克服し、高速性と高精度性を有する UUC 手法を開発することを目的とした。

3. 研究の方法

(1) はじめに先行研究をサーベイし、そこで対象としている問題設定やそれに対して提案されている原理、理論的に予想される利点と問題点、検証計算機実験結果に基づく妥当性と問題点の抽出を行った。

(2) 次に(1)の結果を受けて、現実的に様々なデータ解析で求められる適用条件を踏まえつつ、2の研究の目的を達成するために取り組むべき問題設定を定式化し、目的の達成に向けた新しい UUC 手法の理論構築を行った。

(3) さらに構築した理論に基づき UUC 手法をアルゴリズム化し、それをさらに実際に計算機で実行可能な Python プログラムに実装した。

(4) 上記の問題設定において仮定した諸条件を満たすデータを様々なパラメータ条件で人工的に作成し、それらのデータに上記プログラムを適用して提案 UUC 手法の基本性能評価を行った。さらに機械学習研究分野において標準的に用いられている様々な実ベンチマークデータに適用し、提案 UUC 手法の実用的性能評価を行った。そして、これらの性能評価を通じて提案 UUC 手法が実用的な性能を発揮可能な適用条件や、それが理論予想と整合しているかを確認した。

4. 研究成果

(1) UUC の問題設定 [du Plessis13]

d 次元の特徴空間 $X \subset R^d$ において、事例を $x \in X$ 、正負クラスのラベルを $y \in \{1, -1\}$ とし、クラス割合の異なる2つのラベルなし事例集合を $U_1 = \{x_i\}_{i=1}^n$ と $U_2 = \{x_j\}_{j=1}^{n'}$ が与えられた時、これらの2つの事例集合 U_1 と U_2 それぞれの母集団分布の確率は以下の式で与えられるものとする。

$$\begin{aligned} p_{U_1}(x) &= \sum_{y \in \{1, -1\}} p_{U_1}(x, y) = \sum_{y \in \{1, -1\}} p_{U_1}(x|y) p_{U_1}(y) \\ p_{U_2}(x) &= \sum_{y \in \{1, -1\}} p_{U_2}(x, y) = \sum_{y \in \{1, -1\}} p_{U_2}(x|y) p_{U_2}(y) \end{aligned}$$

ここでは、正負例の周辺分布は U_1 と U_2 に共通で、 $p_{U_1}(x|y) = p_{U_2}(x|y)$ であるとする。また、 U_1 と U_2 のクラス割合は不明であり、互いに異なるものとする。すなわち、 U_1 および U_2 のクラス事前分布 $p_{U_1}(y), p_{U_2}(y)$ は不明かつ $p_{U_1}(y) \neq p_{U_2}(y)$ である。さらに、テストに用いるデータ U_t は正例と負例のクラス事前確率は等しく、 $p_{U_t}(y = 1) = p_{U_t}(y = -1) = 1/2$ とする。また、 U_t の正負例の周辺分布も U_1, U_2 と等しく、 $p_{U_t}(x|y) = p_{U_1}(x|y) = p_{U_2}(x|y)$ とする。

ここでの問題は U_1 と U_2 を学習データとし、新たに与えられた事例のクラスラベルが $y = 1$ か $y = -1$ であるかを判別する分類器 C を学習することである。

(2) 提案 UUC 手法の基本原理

データ U_t に含まれる事例 x のラベルが $y = 1$ であるクラスの事後確率 $p_{U_t}(y = 1|x)$ は、ベイズの定理を用いて

$$p_{U_t}(y = 1|x) = p_{U_t}(x|y = 1)p_{U_t}(y = 1)/p_{U_t}(x)$$

と表すことができる。また $p_{U_t}(y = 1|x)$ と $p_{U_t}(y = -1|x)$ の大小関係により、 $p_{U_t}(y = 1|x) > p_{U_t}(y = -1|x)$ の時は事例 x を $y = 1$ と分類し、 $p_{U_t}(y = 1|x) < p_{U_t}(y = -1|x)$ の時は事例 x を $y = -1$ と分類するという分類基準を用いれば、クラスラベルを定義する式 $d(x)$ は

$$d(x) = \text{sign}[p_{U_t}(y = 1|x) - p_{U_t}(y = -1|x)]$$

とおける。この時、(1) で述べたように仮定として正例と負例のクラス事前確率は等しいとする。つまり

$$p_{U_t}(y = 1) = p_{U_t}(y = -1) = 1/2$$

とすると、クラス事後確率の差は

$$\begin{aligned} & p_{U_t}(y = 1|x) - p_{U_t}(y = -1|x) \\ &= p_{U_t}(x|y = 1)p_{U_t}(y = 1)/p_{U_t}(x) - p_{U_t}(x|y = -1)p_{U_t}(y = -1)/p_{U_t}(x) \\ &= 1/(2p_{U_t}(x))\{p_{U_t}(x|y = 1) - p_{U_t}(x|y = -1)\} \end{aligned}$$

である。以上の式と $1/(2p_{U_t}(x)) > 0$ から、クラスラベルを導く式 $d(x)$ は

$$d(x) = \text{sign}[p_{U_t}(x|y = 1) - p_{U_t}(x|y = -1)]$$

と表せる。

さらに、2つのラベルなし事例集合 U_1 と U_2 の確率密度関数 $p_{U_1}(x)$ と $p_{U_2}(x)$ は

$$\begin{aligned} p_{U_1}(x) &= p_{U_1}(y = 1)p_{U_1}(x|y = 1) + (1 - p_{U_1}(y = 1))p_{U_1}(x|y = -1) \\ p_{U_2}(x) &= p_{U_2}(y = 1)p_{U_2}(x|y = 1) + (1 - p_{U_2}(y = 1))p_{U_2}(x|y = -1) \end{aligned}$$

とおくことができ、2つの確率密度関数の差は

$$\begin{aligned} p_{U_1}(x) - p_{U_2}(x) &= p_{U_1}(y = 1)p_{U_1}(x|y = 1) + (1 - p_{U_1}(y = 1))p_{U_1}(x|y = -1) \\ &\quad - p_{U_2}(y = 1)p_{U_2}(x|y = 1) + (1 - p_{U_2}(y = 1))p_{U_2}(x|y = -1) \end{aligned}$$

と表され、 $p_{U_1}(x|y) = p_{U_2}(x|y)$ を適用すると

$$\begin{aligned} p_{U_1}(x) - p_{U_2}(x) &= p_{U_1}(y = 1)p_{U_1}(x|y = 1) + (1 - p_{U_1}(y = 1))p_{U_1}(x|y = -1) \\ &\quad - p_{U_2}(y = 1)p_{U_1}(x|y = 1) + (1 - p_{U_2}(y = 1))p_{U_1}(x|y = -1) \\ &= (p_{U_1}(y = 1) - p_{U_2}(y = 1))(p_{U_1}(x|y = 1) - p_{U_1}(x|y = -1)) \end{aligned}$$

と表せる。式より、 $p_{U_1}(y = 1) > p_{U_2}(y = 1)$ 、 $p_{U_t}(x|y) = p_{U_1}(x|y) = p_{U_2}(x|y)$ を用いて、クラスラベルを求める式 $d(x)$ は

$$d(x) = \text{sign}[p_{U_1}(x) - p_{U_2}(x)]$$

と表すことができる。すなわち UUC では、正例の割合がより多い事例集合が従う確率密度から正例の割合がより少ない事例集合が従う確率密度の差を求めることで、通常の教師ラベル有り学習の場合と同じクラスラベルの判定を行うことができる。さらに式より、クラスラベルは事例集合 U_1 と U_2 の確率密度の差のみを用いて推定するので、事例集合の正負例の割合はクラスラベルの推定に影響しないことが分かる。

一方で、特徴空間 X 上の事例集合 U_1 または U_2 から選んできた事例 x を、なるべく高精度に事例集合 U_1 と U_2 の何れかに属するかを分類する問題を考える。この場合、 $p_{U_1}(x)$ と $p_{U_2}(x)$ を比べ $p_{U_1}(x) > p_{U_2}(x)$ であるならば事例 x は事例集合 U_1 から抽出された確率の方が高く、 $p_{U_1}(x) < p_{U_2}(x)$ であるならば事例 x は事例集合 U_2 から抽出された確率の方が高いと判断できる。すなわち図 1 に示すように、事例 x が事例集合 U_1 と U_2 の何れに属するかを精度良く推定することは、 $p_{U_1}(x)$ と $p_{U_2}(x)$ の大小関係を推定することと同じであるので、特徴空間 X 上での事例集合 U_1 と U_2 の高精度な分類境界面と $p_{U_1}(x)$ と $p_{U_2}(x)$ の大小関係が反転する境界面は一致する。以上のことから、事例 x の事例集合 U_1 と U_2 各々の確率密度の大小関係を求めることは、事例 x が U_1 に属するか U_2 に属するかを分類することと一致する。さらに、この分類器は U_1 に属すると判定された場合は正例、 U_2 に属すると判定された場合は負例と判定することで、教師ラベル有り分類と等価な正負例の分類を行うことができる。

以上の議論は、任意の高精度な分類モデルや学習アルゴリズムについて成立する。従って、本研究で提案する UUC 手法は、正負例の混合割合が異なる 2つの事例集合 U_1 と U_2 から教師ラベル有り学習と等価な分類器を得るために、正例を含む割合が高い事例集合の方を正例、そうでない事例集合の方を負例として、一般の機械学習や深層学習を適用して分類器学習を行うという、極めて一般的でシンプルなものである。

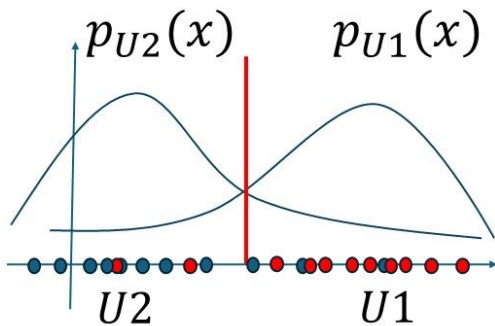


図 1 $p_{U_1}(x), p_{U_2}(x)$ の大小境界面と U_1, U_2 の分類境界面の等価性

(3) 提案 UUC 手法の性能検証結果

○人工データを用いた UUC 手法の評価実験結果

検証に用いる人工データは、正例のクラスラベルを $y = 1$ 、負例のクラスラベルを $y = -1$ とし、正例と負例は各々平均 $\mu^+ = (1,0)$ 、 $\mu^- = (-1,0)$ で分散 $\Sigma = I$ (単位行列) の正規分布に従うものとした。学習データセットに用いる事例集合 $U1$ と $U2$ の確率密度関数 $p_{U1}(x)$ と $p_{U2}(x)$ は、事例集合 $U1$ と $U2$ の正例の割合を各々 q_1 、 q_2 とすると、

$$p_{U1}(x) = q_1 N(\mu^+, I) + (1 - q_1) N(\mu^-, I)$$

$$p_{U2}(x) = q_2 N(\mu^+, I) + (1 - q_2) N(\mu^-, I)$$

と表せる。そこで、事例集合 $U1$ と $U2$ の確率密度関数と各事例数 N_{train} を用いて、事例集合 $U1$ と $U2$ を

$$U1 = \{x_i \sim p_{U1}(x) | i = 1, \dots, N_{train}\}$$

$$U2 = \{x_j \sim p_{U2}(x) | j = 1, \dots, N_{train}\}$$

により作成した。また、テストデータセット Ut は正例と負例のクラス事前確率は等しい ($q = 1/2$) ものとし、その確率密度関数を

$$p_{Ut}(x) = 1/2 N(\mu^+, I) + 1/2 N(\mu^-, I)$$

とした。そして、テストデータセット Ut の事例数を N_{test} として

$$Ut = \{x_\ell \sim p_{Ut}(x) | \ell = 1, \dots, N_{test}\}$$

により作成した。ここで正例の割合 q_1 、 q_2 を変化させて得た事例集合 $U1$ と $U2$ に、新たなクラスラベルとして、 $U1$ には $y = 1$ 、 $U2$ には $y = 0$ を付与した。そして、様々な正例の割合 q_1 、 q_2 を有するデータ $U1$ と $U2$ を用いて分類器を UUC 学習し、テストデータセット Ut を用いて各分類器の性能を検証した。UUC 学習手法としては、線形判別分析とランダムフォレスト、勾配ブースティングの3種類を用いた。

表 1 にこれら 2 つの学習手法による性能比較結果を示す。 $U1(6,4)$ 、 $U2(4,6)$ は、各々の事例集合が正例 6 割と負例 4 割、正例 4 割と負例 6 割を含むことを表す。DSDD は先行研究[du Plessis13] のカーネル法を用いた UUC 手法の結果である。この表が示すように、提案 UUC 手法は学習手法に依存せず従来の UUC 手法よりも遥かに良い結果を示す。 $U1$ 、 $U2$ の他の正負例の混合割合でも (5,5) 以外の場合には同様な結果が得られた。このように、提案 UUC 手法は幅広い学習手法について高性能な分類器をもたらすことが分かった。

表 1 $U1(6,4)$ 、 $U2(4,6)$ の場合の比較結果

	精度	正例に関する F 値	負例に関する F 値
DSDD	0.524 ± 0.035	0.338 ± 0.050	0.627 ± 0.035
線形判別分析	0.835 ± 0.003	0.840 ± 0.002	0.830 ± 0.006
ランダムフォレスト	0.836 ± 0.002	0.837 ± 0.005	0.834 ± 0.001
勾配ブースティング	0.811 ± 0.021	0.816 ± 0.020	0.805 ± 0.021

○ベンチマーク実データを用いた UUC 手法の評価実験結果

機械学習アルゴリズムの実験や学習のためのベンチマークとして広く利用されている UCI (University of California, Irvine) データセット [] から、学習データとして 2 つを選び提案手法性能の比較実験を行った。データセット 1 つ目は、乾燥した豆の品種に関するデータセット (Dry Beans Dataset) を用いた。これは、特徴量として豆の産地や縦横のサイズ、周囲などの 16 次元の数値事例、目的変数として 7 種類の豆の品種の名前で構成されており、事例数は 13611 個のデータセットである。本実験では 2 値分類を行うため、学習データセット中の豆の種類の中で事例数の多い上位 2 種である Deramason 種と Sira 種を正例と負例として扱い、その他の事例は除いた 6182 個のデータセットを用いた。2 つ目は、顧客が定期預金への申し込みをしたか否かに関する銀行のデータセット (Bank Marketing Dataset) を用いた。これは、特徴量として年齢や職業、既婚か否か、ローンの有無などの 7 次元の数値事例と 9 次元の記号事例、目的変数として定期預金に申し込んだか否かで構成されており、事例数は 45211 個のデータセットである。本検証実験では、これらのデータセットを用いて 10 分割交差検証を行った。

図 2 に提案 UUC 手法をランダムフォレスト、勾配ブースティングの学習手法で用いてこれら 2 つのベンチマーク実データに適用した結果と、同じく教師ラベル有り学習手法をこれらの学習手法で用いて同 2 つのデータに適用した結果の比較を示す。(1) は Dry Beans Dataset に適用した結果、(2) は Bank Marketing Dataset に適用した結果である。各々で提案 UUC 手法では正例：負例の割合が 10:0 の $U1$ と 2:8 の $U2$ を用いた場合と同じく 10:0 の $U1$ と 9:1 の $U2$ を用いた場合を示している。教師ラベル有り学習手法については、各々 $U1$ と $U2$ に含まれる正例と負例を取りまとめてラベル有りデータを作成し分類器学習した結果である。提案 UUC 手法は何れの場合も高い分類精度を示している。これに対して教師ラベル有り学習手法では、10:0 の $U1$ と 2:8 の $U2$ をまとめて正例：負例 12:8 の教師ラベル有りデータとして学習した場合は高い精度を

示しているが、10:0 の U1 と 9:1 の U2 をまとめて正例：負例 19:1 の教師ラベル有りデータとして学習した場合は、正例数と負例数が不均衡となるため精度が著しく低下することが分かる。このように、提案 UUC 手法はその理論的性質からラベル情報を用いる教師ラベル有り学習よりも、正例と負例の割合の不均衡にロバストであり、かつ教師ラベル有り学習と同等かそれ以上の分類精度をもたらすという優れた結果を確認した。

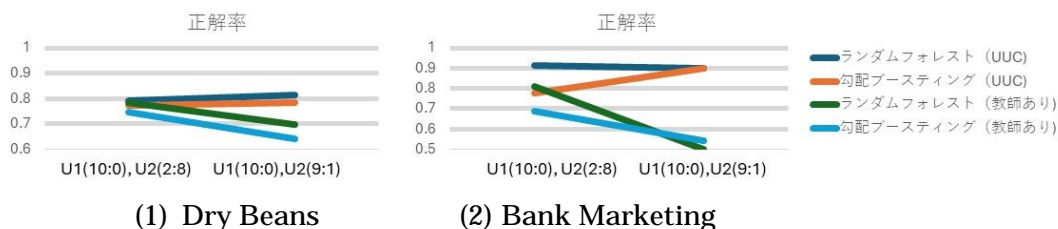


図 2 ベンチマーク実データを用いた評価実験結果

以上のように提案 UUC 手法は、先行研究に比べてシンプルで広範な分類学習手法に適用可能な汎用性を有し、かつ理論的特性と実評価結果の両方で高精度であることが確認された。さらに教師情報を用いる教師ラベル有り学習手法に比べての有利な性能特性を持つことが確認された。

< 引用文献 >

[du Plessis13] M.C. du Plessis, G. Niu and M. Sugiyama, Clustering Unclustered Data: Unsupervised Binary Labeling of Two Datasets Having Different Class Balances, Proc. 2013 Technologies and Applications of Artificial Intelligence, pp.1-6, 2013.

[Kaji19] H. Kaji and M. Sugiyama, Binary Classification Only from Unlabeled Data by Iterative Unlabeled-unlabeled Classification, Proc. ICASSP2019, pp.3527-3531, 2019.

5. 主な発表論文等

〔雑誌論文〕 計7件（うち査読付論文 7件/うち国際共著 5件/うちオープンアクセス 7件）

1. 著者名 Kai Ming Ting, Takashi Washio, Jonathan Wells, Hang Zhang, Ye Zhu	4. 巻 65
2. 論文標題 Isolation Kernel Estimators	5. 発行年 2022年
3. 雑誌名 Knowledge and Information Systems (KAIS Journal)	6. 最初と最後の頁 759-787
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s10115-022-01765-7	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する
1. 著者名 Masaru Kondo, H.D.P. Wathsala, Mohamed S.H. Salem, Kazunori Ishikawa, Satoshi Hara, Takayuki Takaai, Takashi Washio, Hiroaki Sasai and Shinobu Takizawa	4. 巻 5
2. 論文標題 Bayesian optimization-driven parallel-screening on multi-parameters of micromixer-type and organocatalytic conditions in the flowbiaryl synthesis	5. 発行年 2022年
3. 雑誌名 Communications Chemistry	6. 最初と最後の頁 148
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s42004-022-00764-7	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Kai Ming Ting, Jonathan R. Wells, and Takashi Washio	4. 巻 35
2. 論文標題 Isolation Kernel: The X Factor in Efficient and Effective Large Scale Online Kernel Learning	5. 発行年 2021年
3. 雑誌名 Data Mining and Knowledge Discovery	6. 最初と最後の頁 2282-2312
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s10618-021-00785-1	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する
1. 著者名 Kai Ming Ting, Takashi Washio, Jonathan Wells, and Hang Zhang	4. 巻 1
2. 論文標題 Isolation Kernel Density Estimation	5. 発行年 2021年
3. 雑誌名 IEEE ICDM 2021: IEEE ICDM 2021 21st IEEE International Conference on Data Mining	6. 最初と最後の頁 619-628
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICDM51629.2021.00073	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Kai Ming Ting, Takashi Washio, Bi-Cun Xu, Zhi-Hua Zhou	4. 巻 1
2. 論文標題 Isolation Distributional Kernel: A New Tool for Point & Group Anomaly Detection	5. 発行年 2021年
3. 雑誌名 IEEE Transactions on Knowledge and Data Engineering	6. 最初と最後の頁 1-1
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TKDE.2021.3120277	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Takeshi Yoshida, Takashi Washio, Takahito Ohshiro, Masateru Taniguchi	4. 巻 25
2. 論文標題 Classification from Positive and Unlabeled Data Based on Likelihood Invariance for Measurement	5. 発行年 2021年
3. 雑誌名 Intelligent Data Analysis	6. 最初と最後の頁 57-79
掲載論文のDOI (デジタルオブジェクト識別子) 10.3233/IDA-194980	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Kai Ming Ting, Takashi Washio, Bi-Cun Xu, Zhi-Hua Zhou	4. 巻 1
2. 論文標題 Isolation Distributional Kernel: A new tool for kernel based anomaly detection	5. 発行年 2020年
3. 雑誌名 KDD2020: Knowledge Discovery and Data Mining, 2020	6. 最初と最後の頁 233
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3394486.3403062	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

〔学会発表〕 計13件 (うち招待講演 2件 / うち国際学会 3件)

1. 発表者名 横大路 宗征, 鷺尾 隆, 原 聡, ホーランド マシュー
2. 発表標題 UUCによるClass Imbalance Dataからの高精度分類器学習
3. 学会等名 2023年 IBIS Workshop: 情報論的学習理論ワークショップ
4. 発表年 2023年

1. 発表者名 Khalid Md Imrul, 近藤健, 杉寄晃将, H.D.P. Wathsala, 石川一宣, 原聡, 鷹合孝之, 鷲尾隆, 笹井宏明, 滝澤忍
2. 発表標題 機械学習ベイズ最適化を活用するケチミンの電解合成反応条件最適化
3. 学会等名 日本プロセス化学会2018サマーシンポジウム
4. 発表年 2022年

1. 発表者名 森田 泰之, 福田 光宏, 依田 哲彦, 神田 浩樹, 畑中 吉治, 斎藤 高嶺, 田村 仁志, 安田 祐介, 鷲尾 隆, 中島 悠太, 岩崎 昌子
2. 発表標題 ベイズ最適化を用いたイオン源制御手法の開発
3. 学会等名 第19回日本加速器学会年会
4. 発表年 2022年

1. 発表者名 H.D.P. Wathsala, M. Kondo, M.S. H. Salem, K. Ishikawa, S. Hara, T. Takaai, T. Miyazaki, D. Yamashita, T. Washio, H. Sasai and S. Takizawa
2. 発表標題 Bayesian optimization-assisted multi-parameter screening for laboratory- and industrial-scale syntheses
3. 学会等名 2022年度有機合成化学北陸セミナー
4. 発表年 2022年

1. 発表者名 Takashi Washio
2. 発表標題 Measurement Informatics and Its Application in Science
3. 学会等名 SciX2022: SciX (The Great Scientific eXchange) Conference 2022 (The Federation of Analytical Chemistry and Spectroscopy Societies (FACSS)) (招待講演) (国際学会)
4. 発表年 2022年

1. 発表者名 鷺尾 隆
2. 発表標題 革新的先端計測の方程式：計測 + AI = 計測インフォーマティクス
3. 学会等名 人工知能学会「シンポジウム BigDataDX 2022」（招待講演）
4. 発表年 2022年

1. 発表者名 Takayuki Takaai and Makusu Tsutsui
2. 発表標題 Unsupervised Noise Reduction for Nanochannel Measurement Using Noise2Noise Deep Learning
3. 学会等名 PAKDD 2021 Workshops MLMEIN (国際学会)
4. 発表年 2021年

1. 発表者名 Takeshi Yoshida and Eitaro Shinya
2. 発表標題 Class Prior Probability Estimation Using Density Ratio from Unlabeled and Contaminated Positive Datasets
3. 学会等名 PAKDD 2021 Workshops MLMEIN (国際学会)
4. 発表年 2021年

1. 発表者名 山川 将輝, 鷺尾 隆
2. 発表標題 一対比較データによる非結合ガウス過程回帰手法の提案
3. 学会等名 2021年人工知能学会全国大会
4. 発表年 2021年

1. 発表者名 松本 瑞季, 鷺尾 隆
2. 発表標題 クラス事前確率を用いたラベル無しデータからの分類器学習の性能解析
3. 学会等名 2021年人工知能学会全国大会
4. 発表年 2021年

1. 発表者名 鷹合孝之, 筒井真楠, 鷺尾隆
2. 発表標題 Noise2Noise 深層学習を用いた教師無しノノチャンネル計測ノイズ低減
3. 学会等名 人工知能学会第4回計測インフォマティクス研究会
4. 発表年 2020年

1. 発表者名 吉田剛, 新家英太郎, 鷺尾隆
2. 発表標題 ラベルなし事例集合と負事例混入正事例集合からの密度比を用いたクラス事前確率推定
3. 学会等名 人工知能学会第4回計測インフォマティクス研究会
4. 発表年 2020年

1. 発表者名 松本 瑞季, 鷺尾 隆
2. 発表標題 アンサンブル最近傍距離を用いたラベル無しデータからの分類器学習
3. 学会等名 第34回人工知能学会全国大会(2020)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------