Bitcoinサービス業者と利用者アドレスの種類の推定と評価

松本 寛輝[1,a)]　井垣 秀星[2,b)]　菊池 浩明[2,c)]

概要：暗号資産 Bitcoin は仮名により匿名性を担保している．しかし, 匿名性が高い特徴を生かして違法商品の売買やマネーロンダリングなどの取引に Bitcoin が悪用されることがある．また, 公開されている取引記録からは取引の目的を特定することは困難であり, Bitcoin を用いた取引の実態は明らかになっていない．そこで，本稿では，よく知られたサービス業者 3 種と利用者 4 種の計 7 種類の Bitcoin アドレスに着目し，それらの種類の推定を試みる．対象とするサービス提供者は，マイニングプール業者，ATM 業者，Dark web 運営者，サービス利用者は， ATM 利用者，Dark web 利用者, 交換所利用者，掲示板サイト Bitcointalk 登録者である．各アドレスの取引記録を用いて決定木学習を実施することで Bitcoin アドレスの種類属性を推定し, これらのアドレス間の取引の特徴や総量, 頻度などの実態を明らかにする．

キーワード：Bitcoin，暗号資産

1. はじめに

暗号資産の一種である Bitcoin[1] は 2020 年 1 月 23 日の時点で最も時価総額が高い暗号資産とされている [7]. Bitcoin の特長は, 銀行や国家などの第三者機関を介することなく取引を行えることや資産の差し押さえを行うことが困難であることが挙げられる．また, 取引にはアドレスと呼ばれる仮名を用いるため個人の特定が困難であり, 匿名性が高いと言われている．

一方, 公開されている Bitcoin の取引記録からアドレスの分類を行う先行研究がある．2013 年に Ron ら [2] は Bitcoin の全取引を追跡することで大量の Bitcoin を送金しているアドレスの取引構造に特有のパターンが見られることを示している．2013 年に Meiklejohn ら [3] は取引に利用されたアドレスの取引パターンより, 同一ユーザが管理しているアドレスが識別可能であることを示している．また, アドレスによる匿名性ではプライバシーの保護は十分でないことが複数の先行研究で明らかにされている．2015 年に Dupont ら [4] はアドレスについて取引時刻分布に注目し, ユーザが居住している地域のタイムゾーンを推定することが可能であることを示している．2018 年に永田ら [5] は Bitcoin 取引の送金先アドレス集合に注目し, 最大で 80.5%のアドレスを識別することが可能であることを示している．

Dupont ら, 永田らによる研究結果からはアドレスの識別リスクやアドレス所有者の属性を推定されるリスクがあることを示している．しかし, 実験に使用したアドレスは掲示板サイト Bitcointalk 登録者から収集しており, 一般的に利用されている Bitcoin アドレスに関する情報は考慮されていない．また, 投資目的で利用されているアドレスや違法商品の売買やマネーロンダリングなどに悪用されたアドレスなどで異なることが推測される．

一般に Bitcoin アドレスは仮名のためアドレスを所有している人物・団体を知ることではできない．また, アドレスの取引情報は全て公開されているが, 取引が行われた目的は分からない．そこで本研究では, アドレスの所有者と利用したサービスについて着目し, それらの種類の推定を検討する．我々はオンライン上に公開されているアドレスや ATM を用いて取引を行うことで,Bitcoin アドレスを収集した．よく知られたサービス業者 3 種と利用者 4 種の計 7 種類の Bitcoin アドレス, 計 4,049 アドレスを収集した．アドレス 7 種類の推定には取引の入力・出力アドレス数等の取引に関するパターンを特徴量とした決定木学習を実施する．決定木学習の結果から各アドレスが利用したサービスが推定されるリスクを明らかにし, 推定リスクが高い要因となる取引パターンについて考察する．

2. アドレスデータセット

2.1 7 種類の Bitcoin アドレスの定義

本研究ではオンライン上に掲載されているアドレスや

[1] 明治大学大学院先端数理科学研究科
Nakano, Nakano-ku, Tokyo 164–8525, Japan
[2] 明治大学総合数理学部
Nakano, Nakano-ku, Tokyo 164–8525, Japan
[a)] cs192026@meiji.ac.jp
[b)] ev50516@meiji.ac.jp
[c)] kikn@meiji.ac.jp

ATM 実機を用いて取引を行っているアドレスの収集を行った．2019 年 4 月 1 日から同年 9 月 30 日の 6 ヵ月間において 1 度でも取引を行ったアドレスのみを収集対象とした．各アドレスの取引情報は Blockchain Explorer[8] より取得した．表 1 に収集したアドレス数と取引数の内訳を示す．

表 1 収集した Bitcoin アドレスデータセット

項目名	アドレス数		取引数	収集期間
	業者	ユーザ		
Bitcointalk		2,391	29,638	2019/4/1 ～ 2019/9/30
Bitcoin ATM	3	452	26,849	
Dark web	26	67	35,076	
Exchange		1,012	33,351	
Mining Pool	98		24,876	
合計	4,049		149,790	

表 1 の取引数は項目別に集計している．すなわち, 同一項目の業者, ユーザ間での取引は 1 件として集計している．また, 収集したアドレスからは複数の種類に重複しているものを取り除いている．例えば,Bitcointalk と Exchange(交換所) など 2 つ以上の種類属性を持つアドレスは表 1 の中には含まれていないものとする．

本稿では収集した Bitcoin アドレスの所有者をサービス業者と利用者の 2 つに分ける．サービス業者は Bitcoin を利用した商品の販売やサービスを提供するなど営利目的で使用する．利用者は利用した投資や商品の購入, サービスを利用するために Bitcoin を使用する．

2.2 Bitcointalk

Bitcointalk[9] は Bitcoin を主にした暗号資産に関する情報を交換するためのオンラインフォーラムである．Bitcointalk ユーザは図 1 で示すようなプロフィールページを作成し, 自身が管理を行っている Bitcoin アドレスを記載している．アドレスをプロフィールページに記載する理由として Bitcointalk ユーザがフォーラム内で回答した際の寄付を受け付けるため等が考えられる．

本研究では Bitcointalk のプロフィールページに記載されているアドレスはユーザ自身が管理を行っているアドレスと考え,Bitcointalk ユーザのアドレスと定義する．

2.3 Bitcoin ATM

Bitcoin ATM[10] は Bitcoin を預貯金することができるサービスである．Bitcoin ATM の実機を図 2 に示す．利用者は所有しているアドレスの公開鍵情報 (QR コード) を ATM に入力し, 入金したい金額を現金で入れる．すると Bitcoin ATM のアドレスから利用者のアドレスへ送金が行われる．

本研究では世界中に設置されている Bitcoin ATM のうち設置台数が 600 台を超えたカナダにある実機のアドレス

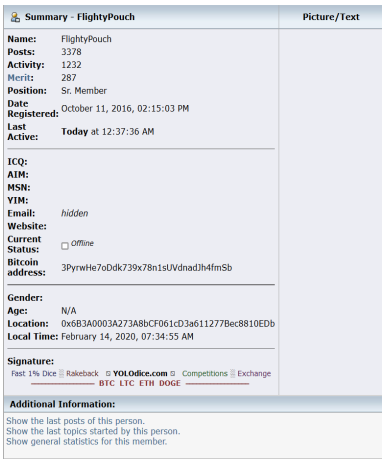

図 1 Bitcointalk プロフィールページ

を収集する．ATM の実機に割り当てられた固定のアドレスは ATM 業者のアドレス,ATM のアドレスから送金されているアドレスを ATM ユーザのアドレスと定義する．

図 2 カナダ トロントに設置されている Bitcoin ATM 実機

2.4 Dark web

Dark web は匿名通信ネットワーク Tor 対応で配信されているウェブページの総称である．Bitcoin アドレスを収集した Dark web サイトの例を図 3 に示す．

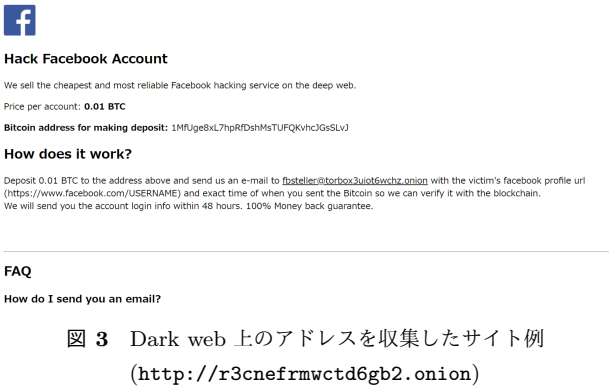

図 3 Dark web 上のアドレスを収集したサイト例 (`http://r3cnefrmwctd6gb2.onion`)

アドレスは Tor ブラウザを利用してアクセス可能な.onion ドメインの web サイトより収集する．本研究ではクレジッ

トカードの売買や SNS ハッキングサービスなど違法性の高いサイトよりサービス利用時に支払い先のアドレスとして指定された Bitcoin アドレスを Dark web 業者と定義する．また,Bitcoin を増額させるサービスを提供していた Dark web のサイト運営者がサービス利用者のアドレスとして公開していた Bitcoin アドレスを Dark web 利用者と定義する．

2.5 Exchange

Exchange(交換所) はユーザが所有している Bitcoin や現金をオンライン上で交換するサービスである．オンライン上で Bitcoin やその他暗号資産の売買を各国の法定通貨を用いて行う．

交換所アドレスは WalletExploer[11] の記載されている Exchanges 一覧リストより 2019 年 4 月 1 日から 2019 年 9 月 30 日まで間に取引が行われたアドレスを収集する．交換所を表 2 に示す．

表 2 アドレスを観測した交換所一覧

交換所名	アドレス数
AnxPro.com	4
BitBay_net	13
Bitstamp_net	40
Bittrex_com	116
CoinHako.com	2
HappyCoins.com	1
Hashnest_com	199
HitBtc_com	89
Kraken_com	26
MercadoBitcoin_com_br	130
OKCoin.com	1
Poloniex_com	110
YoBit_net	281

交換所と取引しているアドレスを交換所ユーザのアドレスと定義する．

2.6 Mining Pool

Mining Pool は Bitcoin の取引情報をまとめたブロックを多数の採掘者 (マイナー) が協力して取引を検証して報酬を得るための仕組みである．ブロックの生成作業はマイニングと呼ばれ, 膨大な計算量が必要である．

本研究では,2019 年 4 月 1 日から同年 9 月 30 日までの間に一度でもマイニング報酬を受け取ったことがあるアドレスをマイニングプール運営者 (業者) が管理しているアドレスと定義する．

3. 提案方式

3.1 取引分析

3.1.1 業者の取引分析

Bitcoin 業者のアドレスに関する取引の統計量を表 3 に,Bitcoin 業者アドレスの取引分布を図 4 に示す．

表 3 Bitcoin 業者の取引数

項目名	平均	最小	中央値	最大	標準偏差
Bitcoin ATM	7,551	111	549	21,993	12,509
Dark web	74	1	6	1,272	250
Mining Pool	271	1	60	4,190	668

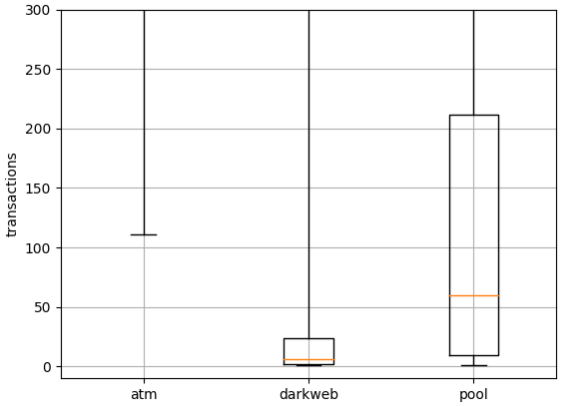

図 4 Bitcoin 業者アドレスの取引分布

表 3 および図 4 について,Bitcoin ATM 業者のアドレス総数は 3 つであることに注意せよ．

3.1.2 ユーザの取引分析

Bitcoin ユーザのアドレスに関する取引の統計量を表 4 に,Bitcoin ユーザアドレスの取引分布を図 5 に示す．

表 4 Bitcoin ユーザの取引数

項目名	平均	最小	中央値	最大	標準偏差
Bitcointalk	13	1	3	722	42
Bitcoin ATM	12	1	2	383	34
Dark web	503	1	23	7,482	1,228
Exchange	45	1	3	4,582	239

図 5 の Bitcoin ユーザの取引分布について,Bitcointalk, ATM, Exchange を利用しているユーザの 75%以上が約 25 回未満の取引を行っていることがわかる．図 4 に示した Bitcoin 業者が管理しているアドレスと比較して取引数が少ないアドレスが多いことが明らかになった．

表 5　取引パターン定義

	入力アドレス数	出力アドレス「おつり」	取引例
$S1$	1	入力アドレスを再指定	基本取引, ATM の「預金」取引
$S2$		入力アドレスを使用しない	一部ウォレットアプリによる取引 (入力アドレスは使い捨て)
$M1$	2 以上	入力アドレスを再指定	交換所のユーザのアドレスを交換所アドレスへまとめる取引
$M2$		入力アドレスを使用しない	マイニングプール業者からのマイナーへ報酬を支払う取引

表 6　取引の特徴量一覧

特徴量	統計量	説明
取引件数	5	取引を行った総回数
送金回数	5	Bitcoin の送金取引を行った総回数
受け取り回数	5	Bitcoin の受け取り取引を行った総回数
取引の入力アドレス数	5	取引時に入力アドレスに使用されたアドレス数
取引の出力アドレス数	5	取引時に出力アドレスに使用されたアドレス数
取引で利用されたアドレス数	1	取引の入力, 出力アドレスに使用されたアドレス数
再利用入力アドレス数	1	異なる取引に繰り返し使用された入力アドレス数
再利用出力アドレス数	1	異なる取引に繰り返し使用された出力アドレス数

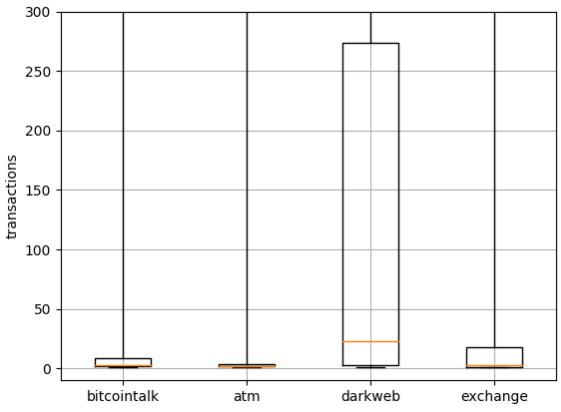

図 5　Bitcoin ユーザアドレスの取引分布

ンがあることを示している [6].

決定木学習で用いる取引のパターンを図 6 および表 5 に示す.

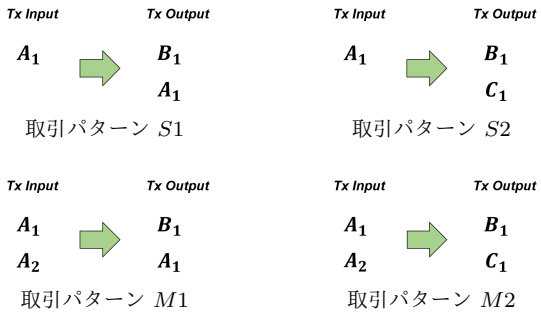

図 6　取引パターン例

3.2　推定手法

3.2.1　決定木学習

本研究では決定木を用いて 7 種類の Bitcoin アドレス種類を推定する. 決定木学習には Python の scikit-learn ライブラリより CART アルゴリズムを使用する. 本実験では 3 クロスバリデーションで交差検証を行った. 学習用・評価用のアドレスはランダムで選択し, 計 100 回実施した推定結果の平均値を求めた. 7 種類のアドレス推定精度は正解率, 適合率, 再現率で評価する.

3.2.2　取引パターンに着目した特徴量

本研究では利用者が取引に用いる入力・出力アドレスの数とおつりを受け取るアドレスに着目した. なぜならば, 取引に用いられるアドレスは利用者が取引を行う際に利用する wallet アプリケーションや利用サービスに依存すると考えられるからである. 例えば,wallet bitpay では Bitcoin の送金を行う際におつりを受け取るアドレスは新たなアドレスを自動で生成する. 我々は ATM 業者と ATM 利用者の取引には特定の入力・出力アドレス数に基づく取引パター

任意の取引を入力アドレスとアドレスの再利用について図 6 で示した次の 4 つのパターンに分類する. パターン $S1$ と $S2$ では入力アドレス数が単一である取引であり, 複数あるもの $M1$ と $M2$ とする. 入力として指定されたアドレスが再び出力 (受け取り) に指定されているパターンを $S1(M1)$ とし, それ以外, すなわち入力と出力に重複はないものを $S2(M2)$ とする.

入力に指定されたアドレスが受け取りに指定される場合, 送金者が取引で生じた「おつり」を受け取っていることを意味している. 例えば, 取引パターン $S1$,$(M1)$ ではアドレス A_1 がおつりを受け取るアドレスとなる. 取引パターン $S2$,$(M2)$ では入力に使用されたアドレスが利用されていない. 本研究では送金者が入力アドレスとは別のアドレスでおつりを受け取っている場合でもおつりを受け取っていない (全金額を送金する) とみなす.

取引パターンに加えて, 表 6 に示した特徴量を用いる. 表 6 の統計量は平均値, 最小値, 中央値, 最大値, 標準偏差の 5 種類である.

表 7　7 種類のアドレスの推定結果

項目名		Bitcoin ATM 業者	Dark web 業者	Mining Pool 業者	Bitcointalk ユーザ	Bitcoin ATM ユーザ	Dark web ユーザ	Exchange ユーザ	計
		予測							
Bitcoin ATM 業者	正解	0	0	0	1	0	0	0	1
Dark web 業者		0	0	0	8	0	0	0	8
Mining Pool 業者		0	0	2	19	8	0	0	29
Bitcointalk ユーザ		0	0	0	633	31	0	53	717
Dark web ユーザ		0	0	0	16	119	0	1	136
Bitcoin ATM ユーザ		0	0	0	12	3	2	3	20
Exchange ユーザ		0	0	0	56	9	0	239	304

4. 実験結果

4.1 取引パターン分析結果

図 6 および表 5 の取引パターンを用いた 7 種類のアドレスの分類結果を表 8 および表 9 に示す.

表 8　7 種類のアドレスの取引パターン数

項目	パターン数			
	S1	S2	M1	M2
Bitcoin ATM 業者	22,319	135	174	25
Dark web 業者	1,242	557	3	127
Mining Pool 業者	19,569	2,845	410	2,052
Bitcointalk ユーザ	6,978	10,704	1,478	10,478
Bitcoin ATM ユーザ	1,700	2,033	44	1,323
Dark web ユーザ	7,627	12,546	1,264	11,711
Exchange ユーザ	8,730	11,269	2,908	10,444

表 9　7 種類のアドレスの取引パターン割合

利用者	パターン [%]			
	S1	S2	M1	M2
Bitcoin ATM 業者	98.5	0.6	0.8	0.1
Dark web 業者	64.4	28.9	0.2	6.6
Mining Pool 業者	78.7	11.4	0.2	6.6
Bitcointalk ユーザ	23.5	36.1	5.0	35.4
Bitcoin ATM ユーザ	33.3	39.9	0.9	25.9
Dark web ユーザ	23.0	37.8	3.8	35.3
Exchange ユーザ	26.2	33.8	8.7	31.3

表 8 の取引数は種類別で集計していることに注意せよ. 表 1 では取引数を項目別に集計しているため, 表 8 では ATM 業者とユーザ間の取引数が多く集計されている.

4.2 推定結果

実験結果を表 10 および実験結果を表 7 に示す. 決定木学習で得られた決定木を図 7 に示す. 図 7 はルートからの深さが 5 以下かつ 1 つの葉に属するサンプルの割合が 10% 以上の特徴量のみ枝刈りして示す. 決定木のルート特徴量 min_input_address とルート特徴量による分類が真の特徴量 mean_output_address の関係を図 8 に示す. 表 6 における特徴量 min_input_address は取引の入力アドレス数の最小値, 特徴量 mean_output_address は取引の出力アドレス数の平均値となる.

表 10　推定実験結果

項目名	正解率		適合率		再現率	
	業者	ユーザ	業者	ユーザ	業者	ユーザ
Bitcointalk		77%		65%		63%
Bitcoin ATM	99%	91%	16%	45%	22%	40%
Dark web	98%	93%	6%	49%	9%	36%
Exchange		85%		80%		79%
Mining Pool	92%		70%		65%	
合計	81%		49%		39%	

各アドレスの推定結果を表 7 に示す. Bitcoin ATM 業者および Dark web 業者はアドレス種類を正しく推定することはできなかった. 最も誤検知が多いアドレスは Bitcointalk ユーザであり,112 アドレスが誤検知となっている. また,Bitcointalk ユーザは 7 種類のアドレスのうち推定率が最も高く約 88%であり, 評価に使用した 717 アドレスのうち 613 のアドレスを正しく推測する.

入力に使用したアドレス数の最小値の特徴量のアドレス分布を図 9 に示す. 入力アドレス数の最小値に関する各アドレスの統計量を表 11 に示す.

4.3 考察

実験結果より, 表 10 の正解率において BitcoinATM 業者・ユーザおよび Dark web 業者・ユーザは 90%以上の値となっている. これは真陽性となるアドレス数に対して真陰性となるアドレス数が多いため高い正解率となっている. そのため, それぞれの適合率および再現率の値は全て 50% 未満となっている.

7 種類のアドレスについて正解率・適合率・再現率全てにおいて約 80%の値となった要素は交換所アドレスであった. また, 交換所のアドレスが推定される要因として取引を行う際に入力に使用するアドレス数の最低値が他のアドレス種類と比較して多いことが要因であることが考えられる.

表 9 の取引パターンについて,Bitcoin ユーザは 4 つのパ

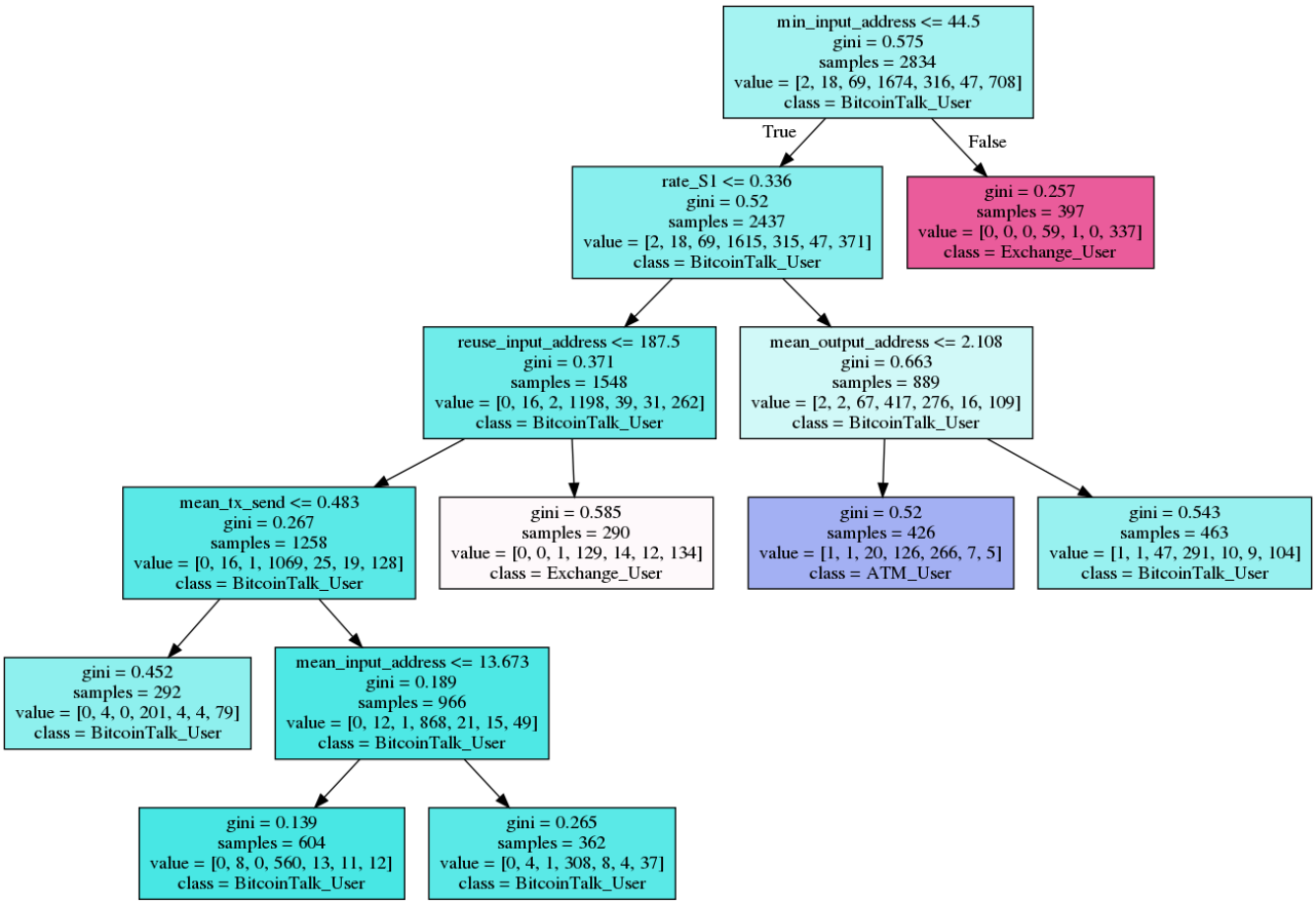

図 7 scikit-learn による決定木モデル例

表 11 7 種類のアドレスの最小入力アドレス数

項目名	平均	最小	中央値	最大	標準偏差
Bitcoin ATM 業者	1	1	1	1	0
Dark web 業者	1.9	1	1	17	3.2
Mining Pool 業者	1	1	1	1	0
Bitcointalk ユーザ	7	1	1	676	40.1
Bitcoin ATM ユーザ	1.3	1	1	112	5.2
Dark web ユーザ	1.7	1	1	12	2.3
Exchange ユーザ	137.9	1	10.5	662	190

ターンに均等な割合となっていた．これは Bitcoin ユーザが行う取引はユーザごとによって複数の異なるサービスを利用していることが原因であると考えられる．業者のアドレスでは取引パターン $S1$ と $S2$ となる割合は約 90 を%超えていた．業者のアドレスはサービスを提供する際に自動化された取引やアドレスを公開することを前提とするため,$S1$ または $S2$ の単一の入力アドレス数が繰り返し利用されていることが考えられる．

実験に使用した特徴量は取引パターンに着目していたが, 学習を行うアドレス数の差やアドレスの取引数に大きく依存してしまい, 推定率が低くなったことが考えられる．本稿では 7 種類のアドレス推定を行ったが, 表 9 の結果を見ると Bitcoin ユーザと業者など分類の分けによっては推定率が変化することが予測される．

5. まとめ

本実験ではよく知られたサービス提供業者の Bitcoin アドレス 3 種類および利用者 4 種類のアドレスを利用し, 決定木学習を用いてアドレス種類の推定を行った．実験結果より最も推定リスクが高い種類は Exchange ユーザのアドレスであり,80%の適合率と 79%の再現率の結果となった．

今回の実験では使用したアドレス種類のデータセットにおいてアドレス数が大きく異なるという問題があるため, アドレス数を増やすまたはアドレス数に影響を受けない推定手法を提案することを今後の課題とする．

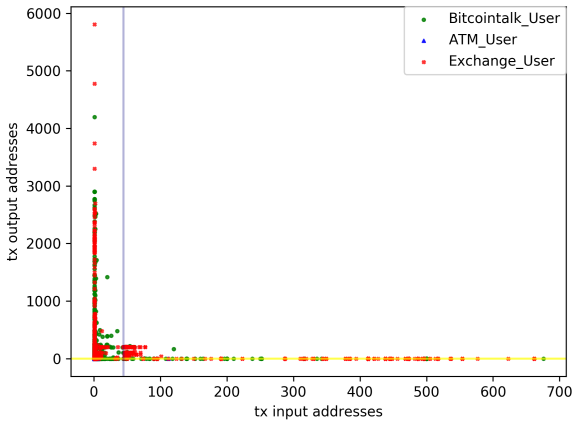

図 8 scikit-learn で作成した決定木の特徴量と推定アドレス分布

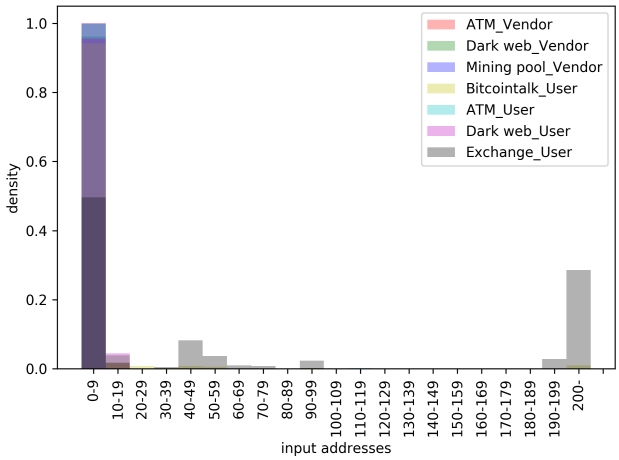

図 9 最小入力アドレス数と 7 種類のアドレスのヒストグラム

参考文献

[1] Satoshi Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System”, `https://bitcoin.org/bitcoin.pdf`
[2] Dorit Ron, Adi Shamir, “Quantitative Analysis of the Full Bitcoin Transaction Graph”, Financial Cryptography and Data Security(FC 2013), pp 6-24, 2013.
[3] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M. Voelker, Stefan Savage, “A fistful of bitcoins: characterizing payments among men with no names”, In Proceedings of the 2013 conference on Internet measurement conference(IMC ’13), Pages 127–140, 2013.
[4] Jules Dupont, Anna C Squicciarini, “Toward De-Anonymizing Bitcoin by Mapping Users Location”, In Proceedings of the 5th ACM Conference on Data and Application Security and Privacy(CODASPY ’15), Pages 139–141, 2015.
[5] 永田倖大, 菊池浩明, “Bitcoin アドレスの送金先集合に基づく匿名性の評価”, 情報処理学会 第 80 回コンピュータセキュリティ研究発表会（CSEC-80）,pp. 1-6, 2018.
[6] 井垣秀星, 松本寛輝, 菊池浩明, “カナダにおける Bitcoin ATM の利用者調査”, 情報処理学会 第 82 回全国大会, 金沢, 2020 年 3 月発表予定.
[7] CoinMarketCap 仮想通貨時価総額上位 100 `https://coinmarketcap.com/ja/` 2020/01/23 7:52 参照
[8] Blockchain Explorer (`https://www.blockchain.com/ja/explorer`)
[9] bitcointalk (`https://bitcointalk.org/`)
[10] Coin ATM Radar Bitcoin ATM Map (`https://coinatmradar.com/`) 2020 年 2 月 14 日 16:32 参照
[11] WalletExplorer.com (`https://www.walletexplorer.com/`)