2020年3月12日

第182回DPS・第88回CSEC合同研究発表会

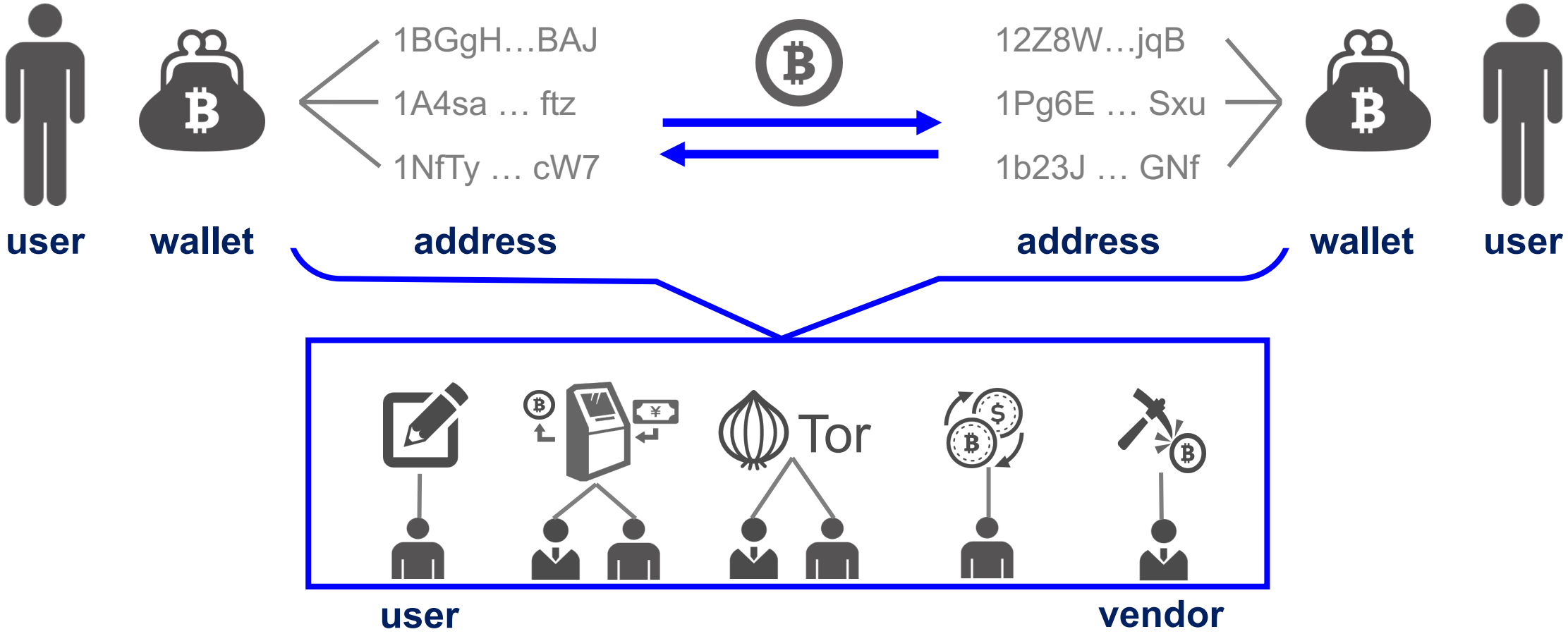
Bitcoinサービス業者と利用者アドレスの種類の推定と評価

松本 寛輝[1]　井垣 秀星[2]　菊池 浩明[2]

1. 明治大学大学院先端数理科学研究科
2. 明治大学総合数理学部

本研究の概要

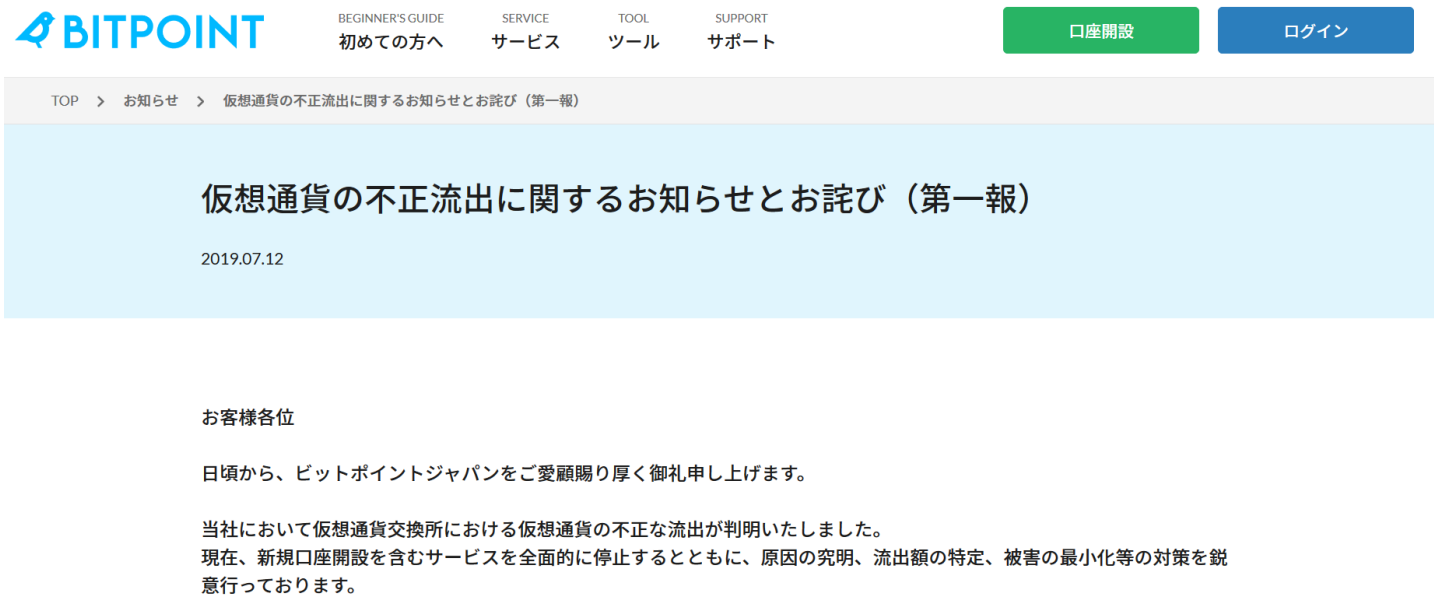
- Bitcoinの取引情報からアドレスに対して7種類の属性を推定
  - **Bitcoinのアドレス種類(属性)推定リスクを明らかにする**



背景

- 暗号資産(仮想通貨)のハッキング被害が後を絶たない
  - 2018年1月 Coincheck（被害額は約580億円)
  - 2018年9月 Zaif（被害額は約70億円)
  - 2019年7月 BITPoint（被害額は約35億円) [1]

- 盗まれた資産の行方
  - 異なる暗号資産への交換[2]
  - 資金洗浄(マネーロンダリング)
  - **資金の利用方法や特定が困難**

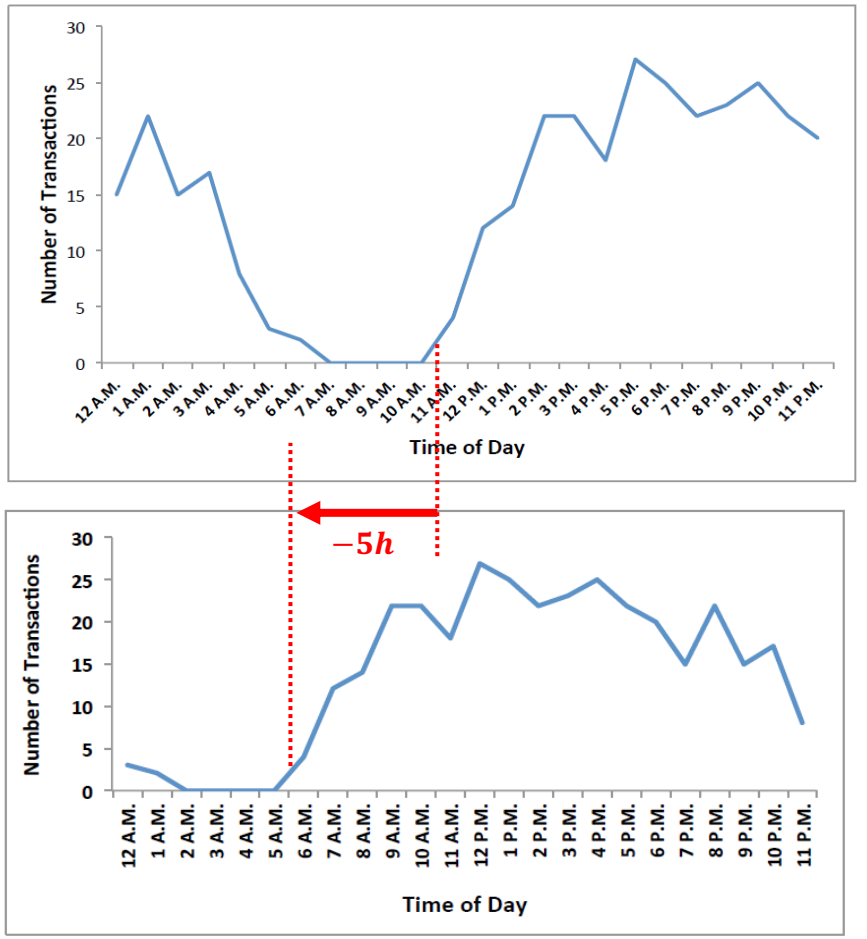


[1] ビットポイントジャパン 仮想通貨の不正流出に関するお知らせとお詫び（第一報）(https://www.bitpoint.co.jp/news/info/info-2019071203/)
[2] 朝日新聞 コインチェック事件、巨額のＮＥＭ交換した男らを立件へ (https://www.asahi.com/articles/ASMDN51V8MDNUTIL01Z.html)

関連研究：Bitcoinと匿名性

- Bitcoinアドレスの識別
  - 同一ユーザが管理するアドレスを識別 (Meiklejohn, 2013)
  - 送金先情報に基づくアドレスの識別 (永田, 2018)

- Bitcoinユーザの属性推定
  - アドレス管理者のタイムゾーンを特定 (Dupont, 2015)
  - Bitcoinユーザのタイムゾーン属性の推定 (井垣, 2019)



Dupontら[3]によるタイムゾーン推定

- 上図 あるユーザの取引時間推移 ($UTC \pm 0$)
- 下図 あるユーザのタイムゾーンに基づく取引時間推移 ($UTC - 5$)

[3] J. Dupont, A. C. Squicciarini. Toward De-Anonymizing Bitcoin by Mapping Users Location. In Proceedings of Conference on Data and Application Security and Privacy (CODASPY'15). ACM, 2015

関連研究：Bitcoinと匿名性

- Bitcoinアドレスの識別
  - 同一ユーザが管理するアドレスを識別 (Meiklejohn, 2013)
  - 送金先情報に基づくアドレスの識別 (永田, 2018)

- Bitcoinユーザの属性推定
  - アドレス管理者のタイムゾーンを特定 (Dupont, 2015)
  - Bitcoinユーザのタイムゾーン属性の推定 (井垣, 2019)

識別・属性推定に利用しているアドレスが主に
Bitcointalk(オンラインフォーラム)の情報

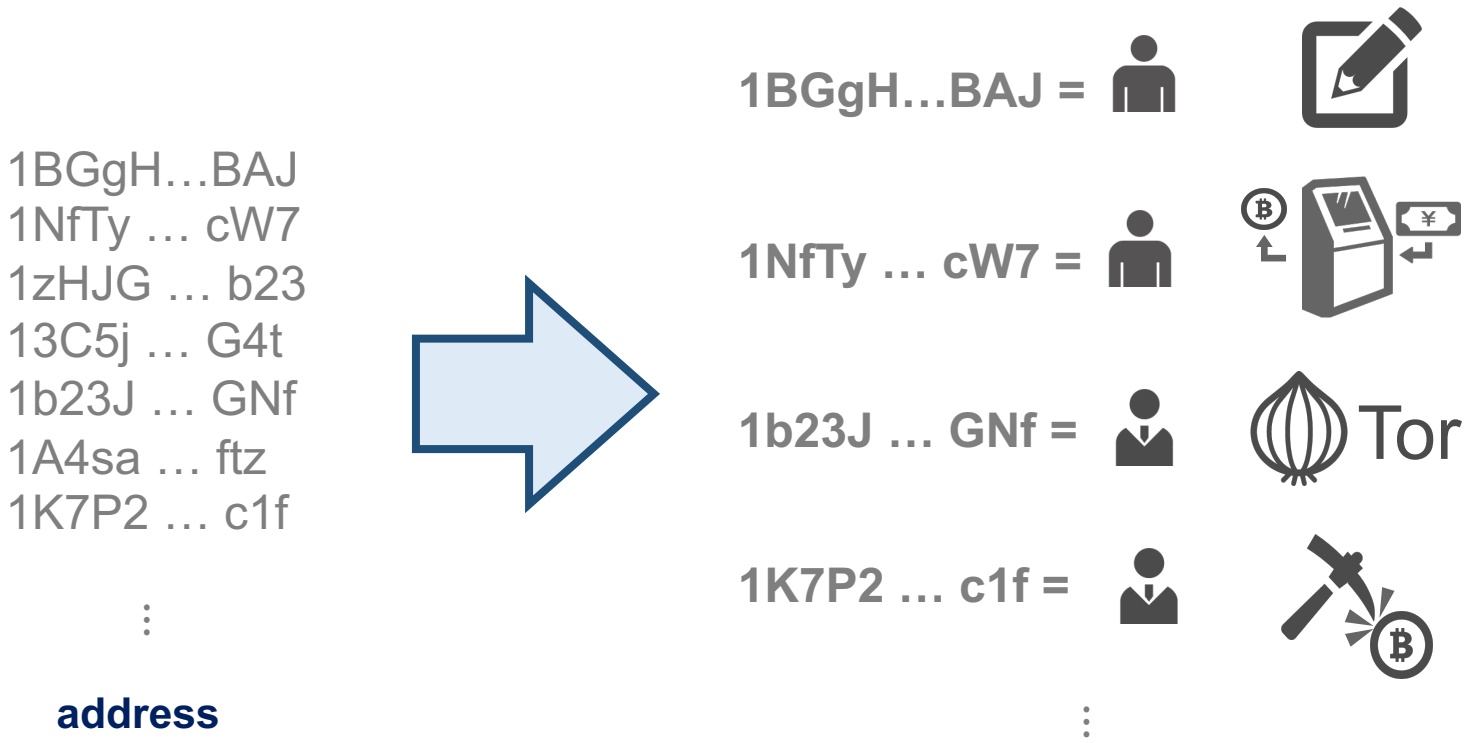
➔ 実験結果・評価に偏りが出る可能性がある

Bitcointalk プロフィールページ例

[3] J. Dupont, A. C. Squicciarini. Toward De-Anonymizing Bitcoin by Mapping Users Location. In Proceedings of Conference on Data and Application Security and Privacy (CODASPY'15). ACM, 2015

研究目的

- Bitcoinの取引情報からアドレス種類を識別する
  - アドレスが利用しているサービスによって識別・推定リスクが異なると予測
  - 識別されやすいアドレスの種類, 特徴を明らかにする



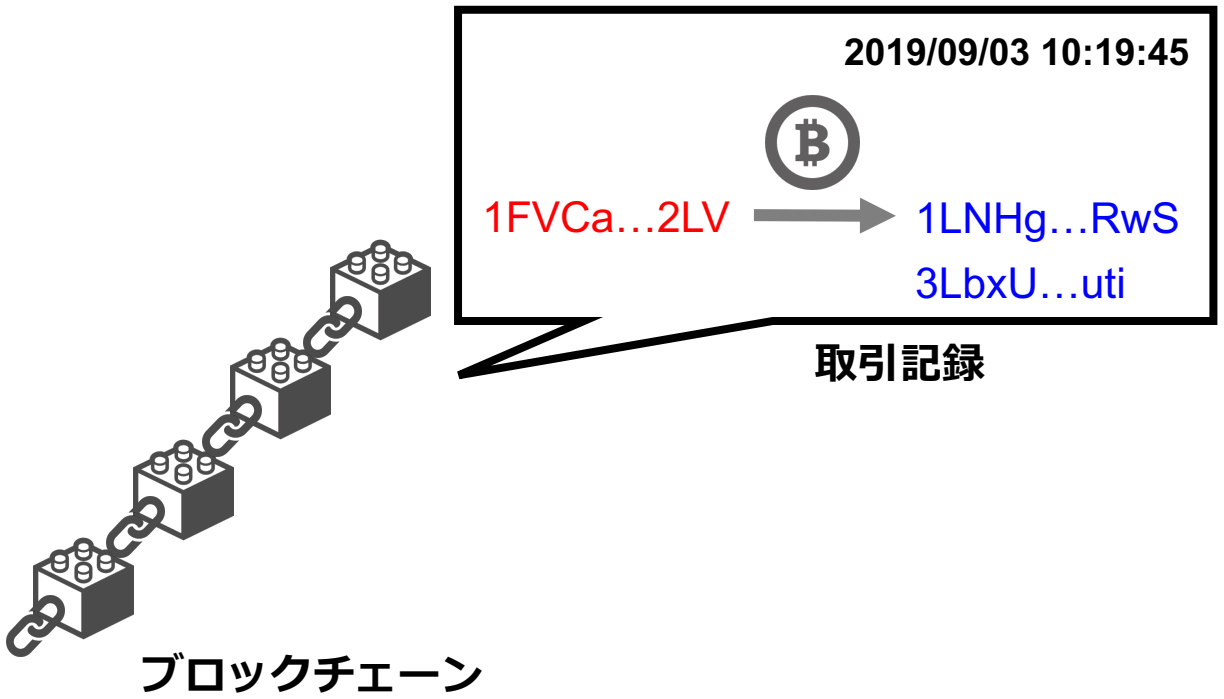
Bitcoinアドレスの利用者を
業者(3種)と**ユーザ(4種)**の
計7種類とした

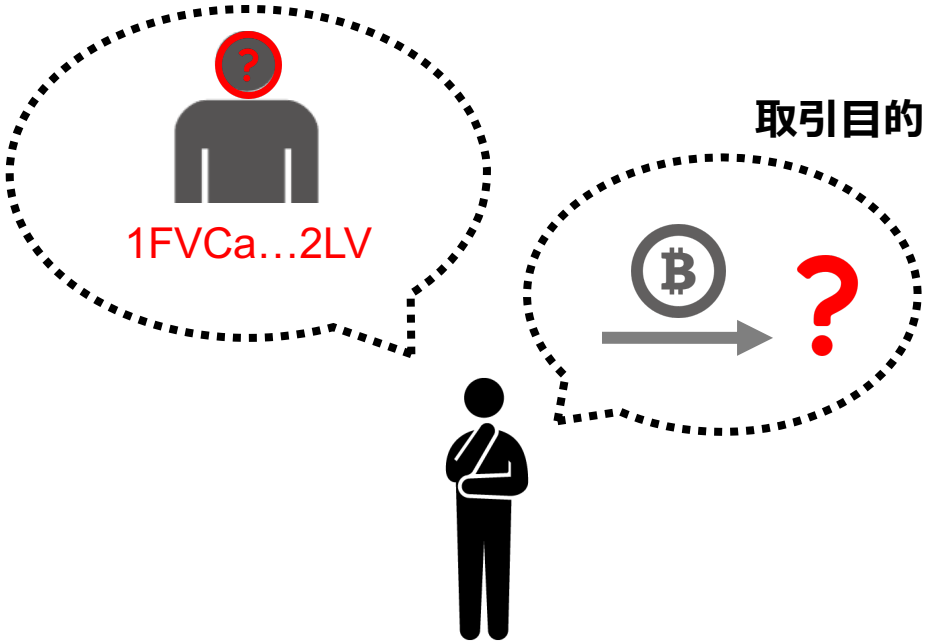
業者(3種)

ユーザ(4種)

問題点：Bitcoinの匿名性

- 正解データがわからない
  - Bitcoinの取引記録はすべて公開されている
  - Bitcoin取引の目的(投資, 買い物等)がわからない





本研究のアプローチ

- 5項目7種類のBitcoinアドレスデータを収集
  - 7種類のアドレスデータに関する取引情報を取得
    - » 複数の種類に重複しているアドレスは除外
  - 取引情報を用いた決定木学習を実施, 7種のアドレス推定を試みる

Bitcointalk (オンラインフォーム)

プロフィール登録している
利用ユーザ

BitcoinATM (カナダ:トロントの実機)

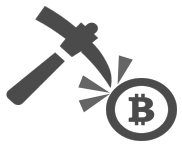
ATMを管理する業者
ATMを利用して預貯金を
行っている利用ユーザ

Darkweb

違法商品・サービスを提供する業者
Darkweb上で公開されている
利用ユーザ

Exchange (交換所)

サービスを利用しているユーザ



MiningPool(マイナー)

マイニング報酬を受け取る業者

データセット(アドレス数・取引数)

- 取引所Blockchain Explorer[4]の取引データよりアドレスの取引情報を取得

項目名	アドレス数		取引数	期間
	業者	ユーザ		
Bitcointalk		2,391	29,638	2019/4/1 ～ 2019/9/30 (6ヶ月間)
BitcoinATM	3	452	26,849	
Darkweb	26	67	35,076	
Exchange		1,012	33,351	
MiningPool	98		24,876	
全体	4,049		149,790	

[4] BLOCKCHAIN.COM (https://www.blockchain.com/explorer)

特徴量：取引パターン

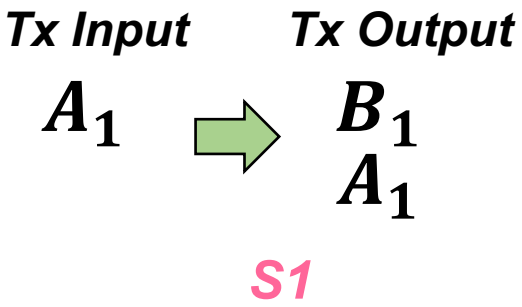
- 取引パターン$(S1, S2, M1, M2)$の定義

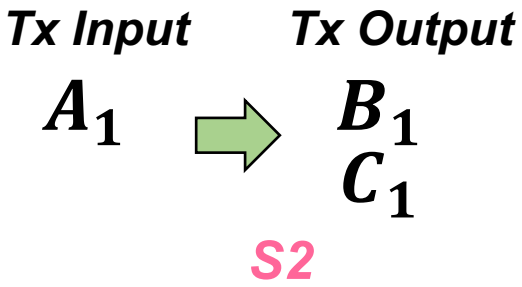
	入力アドレス数が ”1個”	入力アドレス数が ”複数”
入力アドレスと出力アドレスに重複が”ある” ➔ おつりを受け取る	*Tx Input* A_1 ⇒ *Tx Output* B_1, A_1 ***S1***	*Tx Input* A_1, A_2 ⇒ *Tx Output* B_1, A_1 ***M1***
入力アドレスと出力アドレスに重複が”ない” ➔ おつりを受け取らない	*Tx Input* A_1 ⇒ *Tx Output* B_1, C_1 ***S2***	*Tx Input* A_1, A_2 ⇒ *Tx Output* B_1, C_1 ***M2***

特徴量：取引パターン

- 7種類のアドレスの取引パターン割合
  - 業者アドレス$S1, S2$に注目

$S1, S2$ … 入力アドレス数が ”1個”





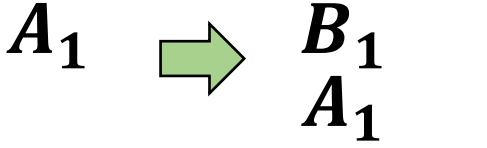
利用者	パターン[%]			
	$S1$	$S2$	$M1$	$M2$
BitcoinATM 業者	98.5	0.6	0.8	0.1
Darkweb 業者	64.4	28.9	0.2	6.6
MiningPool 業者	78.7	11.4	0.2	6.6
Bitcointalk ユーザ	23.5	36.1	5.0	35.4
BitcoinATM ユーザ	33.3	39.9	0.9	25.9
Darkweb ユーザ	23.0	37.8	3.8	35.3
Exchange ユーザ	26.2	33.8	8.7	31.3

特徴量：取引パターン

- 7種類のアドレスの取引パターン割合
  - 業者アドレス$S1, S2$に注目

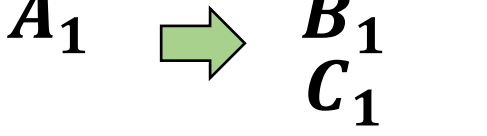
$S1, S2$ … 入力アドレス数が ”1個”

Tx Input *Tx Output*



S1

Tx Input *Tx Output*



S2

業者アドレスが行った取引では
約90%以上が$S1, S2$に分類される

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

特徴量：取引の統計量

- 取引の統計量
  - 統計量には平均値・ 最小値・中央値・最大値・標準偏差の5種類を使用

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	取引の入力アドレス, 出力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

アドレス種類推定：実験概要

- 実験概要
  - 実験目的：7種類のアドレス種類推定
  - 推定方法：決定木学習(scikit-learn, CARTアルゴリズム)を利用
  - 評価指標：アドレス推定結果の正解率・適合率・再現率

- 実験手順
  1. 7種類のアドレスに対して決定木学習を用いて学習・評価を実施
     » 3クロスバリデーションでの交差検証を実施
  2. 評価データ全体と7種類のアドレスに対する正解率・適合率・再現率を記録
  3. 手順1・2を100回繰り返し実施, 正解率・適合率・再現率の平均を算出

アドレス種類推定：実験結果

- アドレスの推定結果
  - 決定木学習による正解率・適合率・再現率を算出

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
BitcoinATM	99	91	16	45	22	40
Darkweb	98	93	6	49	9	36
Exchange		85		80		79
MiningPool	92		70		65	
全体	81		49		39	

アドレス種類推定：実験結果

Exchangeユーザは
正解率・適合率・再現率が
約8割となっている

- アドレスの推定結果
  - 決定木学習による正解率・適合率・再現率を算出

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
BitcoinATM	99	91	16	45	22	40
Darkweb	98	93	6	49	9	36
Exchange		85		80		79
MiningPool	92		70		65	
全体	81		49		39	

7種類のアドレス推定結果

Bitcointalkユーザ
- 誤検知が最も多い(112アドレス)
- 正しく推定したアドレス数が最も多い(約88%)

■ 決定木学習による7種類のアドレスの推定結果の一例

項目名		BitcoinATM 業者	Darkweb 業者	MiningPool 業者	Bitcointalk ユーザ	BitcoinATM ユーザ	Darkweb ユーザ	Exchange ユーザ	合計
		予測							
BitcoinATM 業者	正解	0	0	0	1	0	0	0	1
Darkweb 業者		0	0	0	8	0	0	0	8
MiningPool 業者		0	0	2	19	8	0	0	29
Bitcointalk ユーザ		0	0	0	633	31	0	53	717
BitcoinATM ユーザ		0	0	0	16	119	0	1	136
Darkweb ユーザ		0	0	0	12	3	2	3	20
Exchange ユーザ		0	0	0	56	9	0	239	304

7種類のアドレス推定結果

Exchange(交換所)ユーザ
- Bitcointalkユーザの誤検知・誤推定が最も多い
  ➔ 未知の交換所を利用している可能性

■ 決定木学習による7種類のアドレスの推定結果の一例

項目名		BitcoinATM 業者	Darkweb 業者	MiningPool 業者	Bitcointalk ユーザ	BitcoinATM ユーザ	Darkweb ユーザ	Exchange ユーザ	合計
		予測							
BitcoinATM 業者	正解	0	0	0	1	0	0	0	1
Darkweb 業者		0	0	0	8	0	0	0	8
MiningPool 業者		0	0	2	19	8	0	0	29
Bitcointalk ユーザ		0	0	0	633	31	0	53	717
BitcoinATM ユーザ		0	0	0	16	119	0	1	136
Darkweb ユーザ		0	0	0	12	3	2	3	20
Exchange ユーザ		0	0	0	56	9	0	239	304

決定木学習モデル例

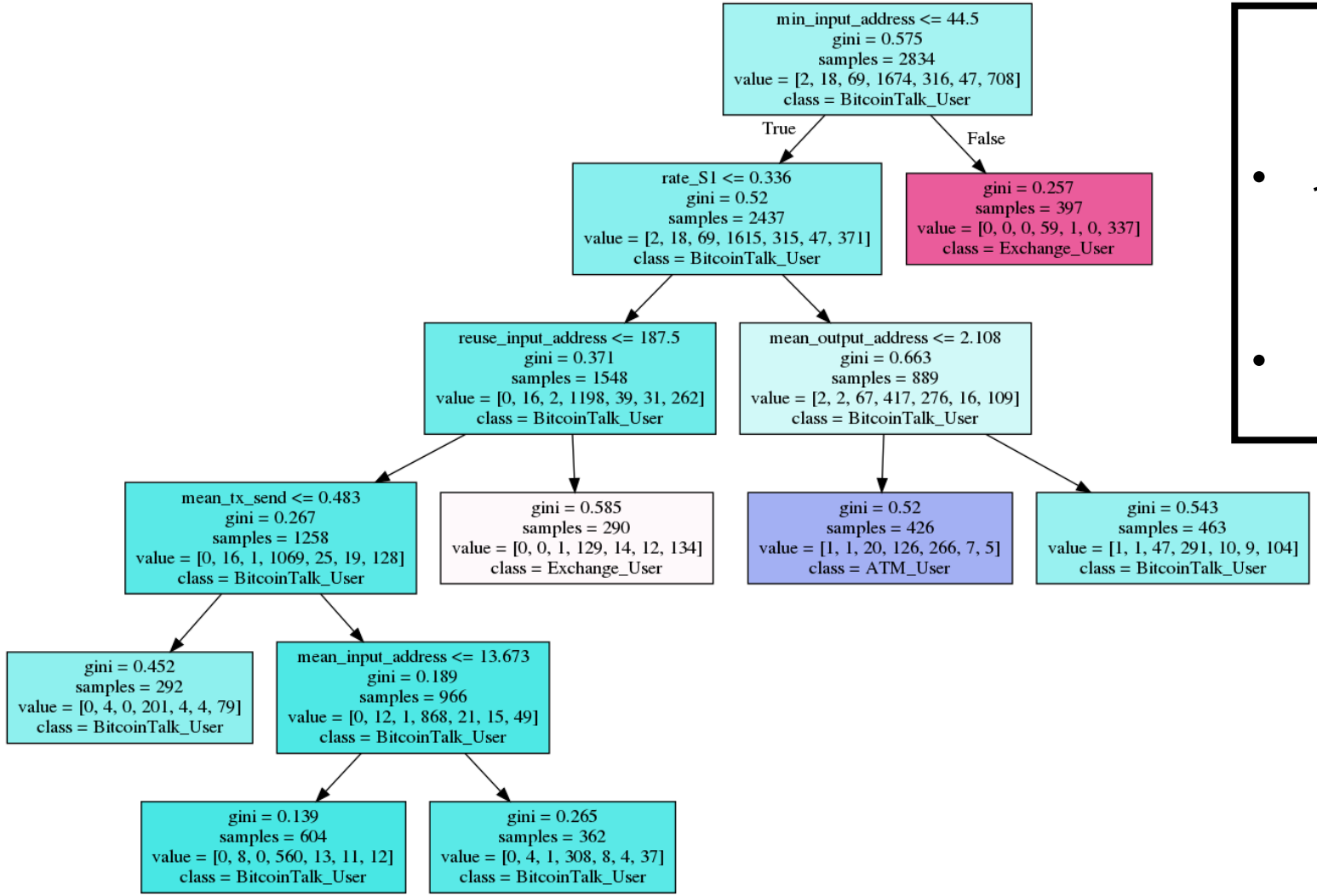


図　作成した決定木モデルの一例

左図の決定木モデルについて

- 以下の条件で枝刈りを実施
  - » ルートからの深さ5以下
  - » 1つの葉に属するサンプルの割合が10%以上
- 100回作成した決定木モデルの一例

決定木学習モデル例：ルート特徴量

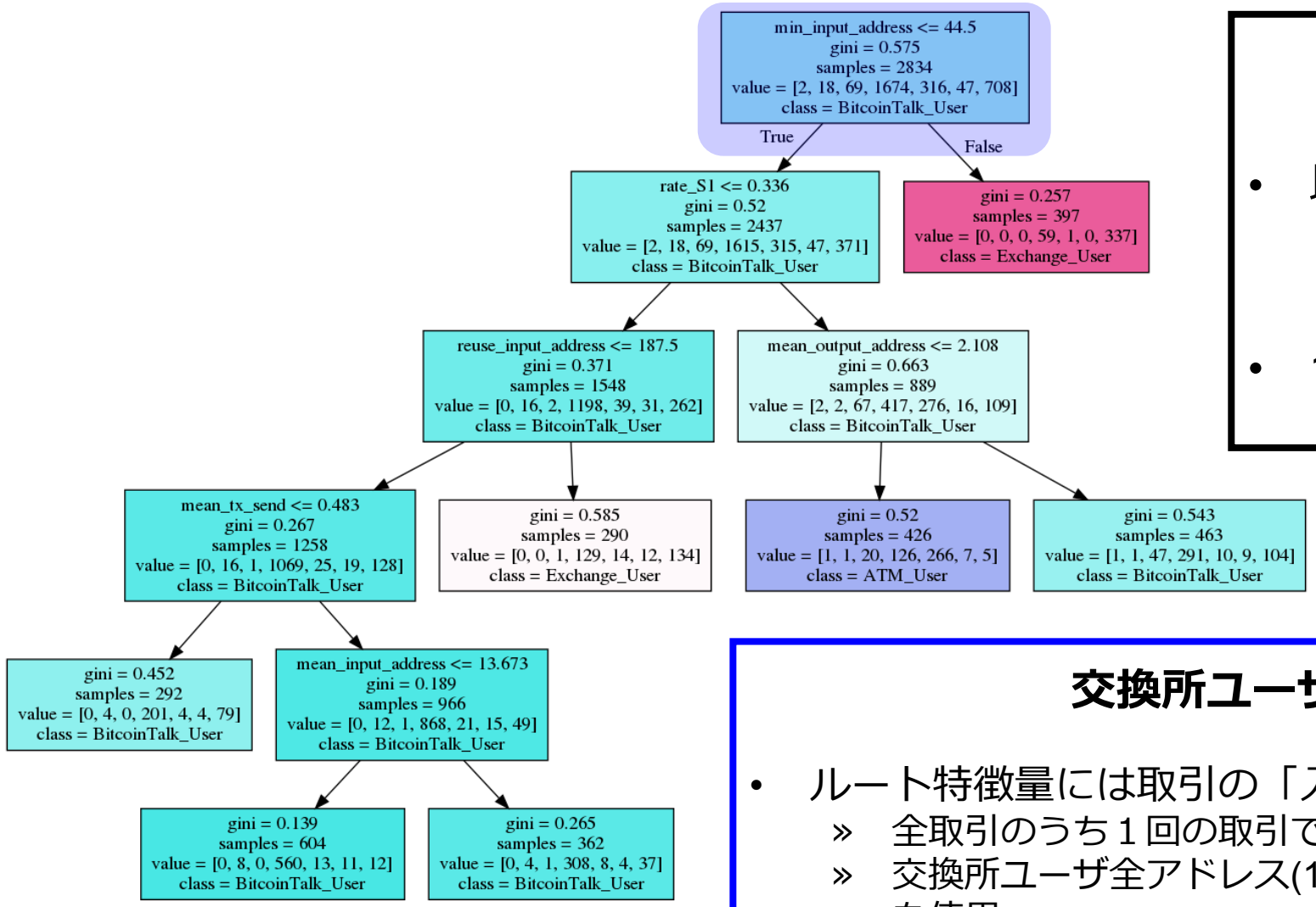


図　作成した決定木モ

左図の決定木モデルについて

- 以下の条件で枝刈りを実施
  - » ルートからの深さ5以下
  - » 1つの葉に属するサンプルの割合が10%以上
- 100回作成した決定木モデルの一例

交換所ユーザの推定(ルート特徴量)

- ルート特徴量には取引の「入力アドレス数の最小値」を利用
  - » 全取引のうち１回の取引で使用した入力アドレス数の最低値
  - » 交換所ユーザ全アドレス(1012個)の約48%が”**最低44.5個以上のアドレス**”を使用

ルート特徴量：入力アドレス数の最小値

表　入力アドレス数の最小値の統計量

利用者	平均	最小	中央値	最大	標準偏差
BitcoinATM 業者	1	1	1	1	0
Darkweb 業者	1.9	1	1	17	3.2
MiningPool 業者	1	1	1	1	0
Bitcointalk ユーザ	7	1	1	676	40.1
BitcoinATM ユーザ	1.3	1	1	112	5.2
Darkweb ユーザ	1.7	1	1	12	2.3
Exchange ユーザ	137.9	1	10.5	662	190

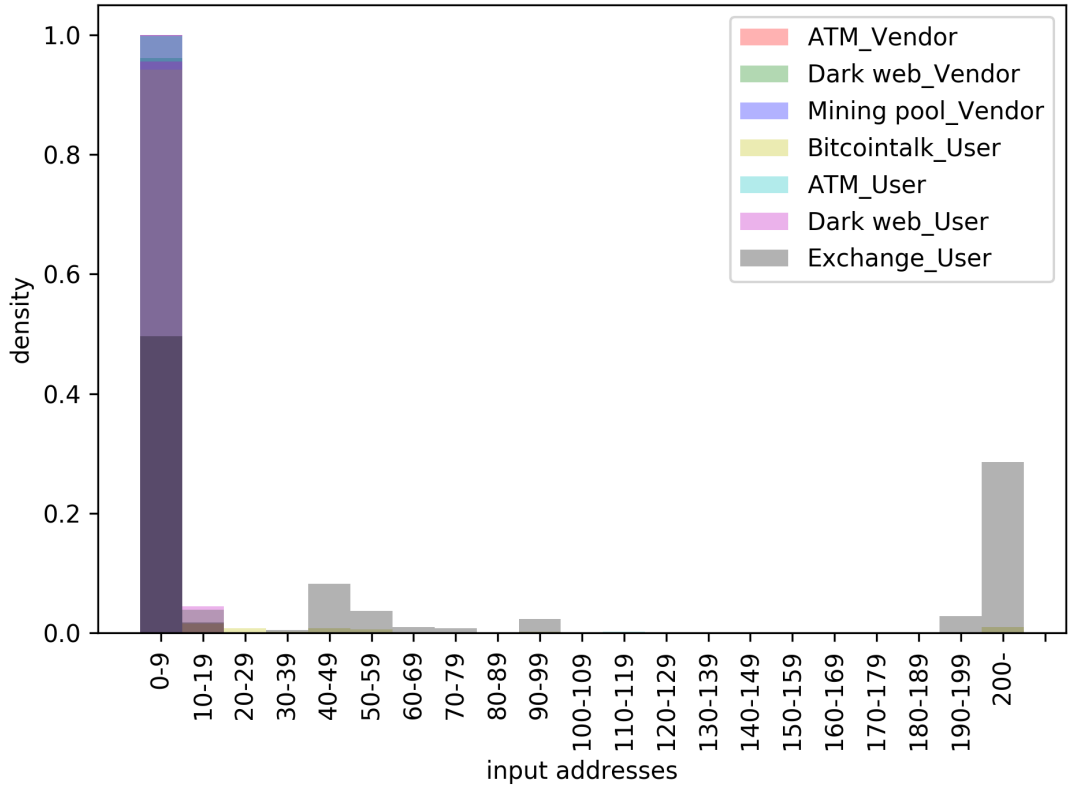


図　入力アドレス数の最小値のヒストグラム

決定木学習モデル例：3種類のアドレス推定

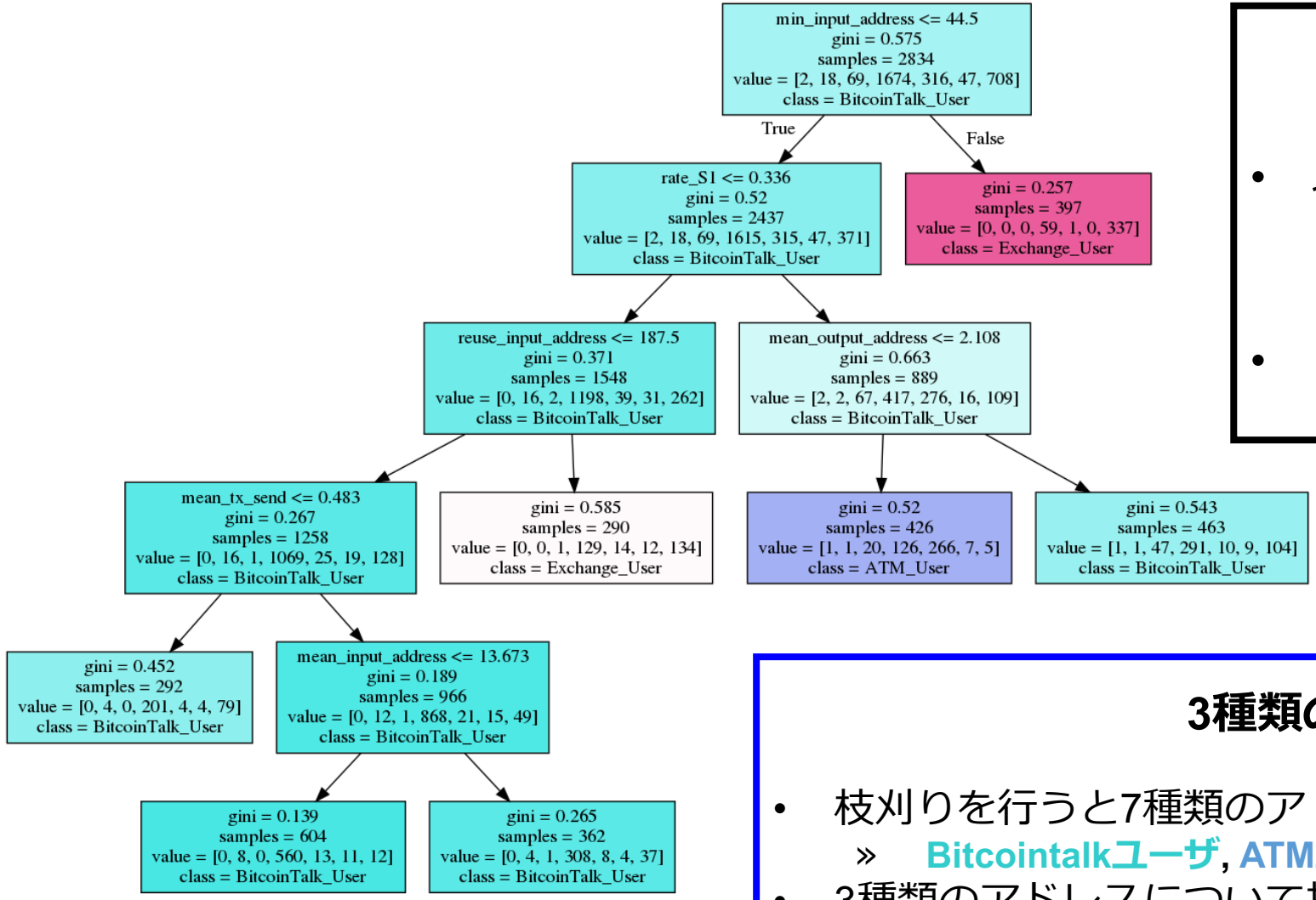


図　作成した決定木モ

左図の決定木モデルについて

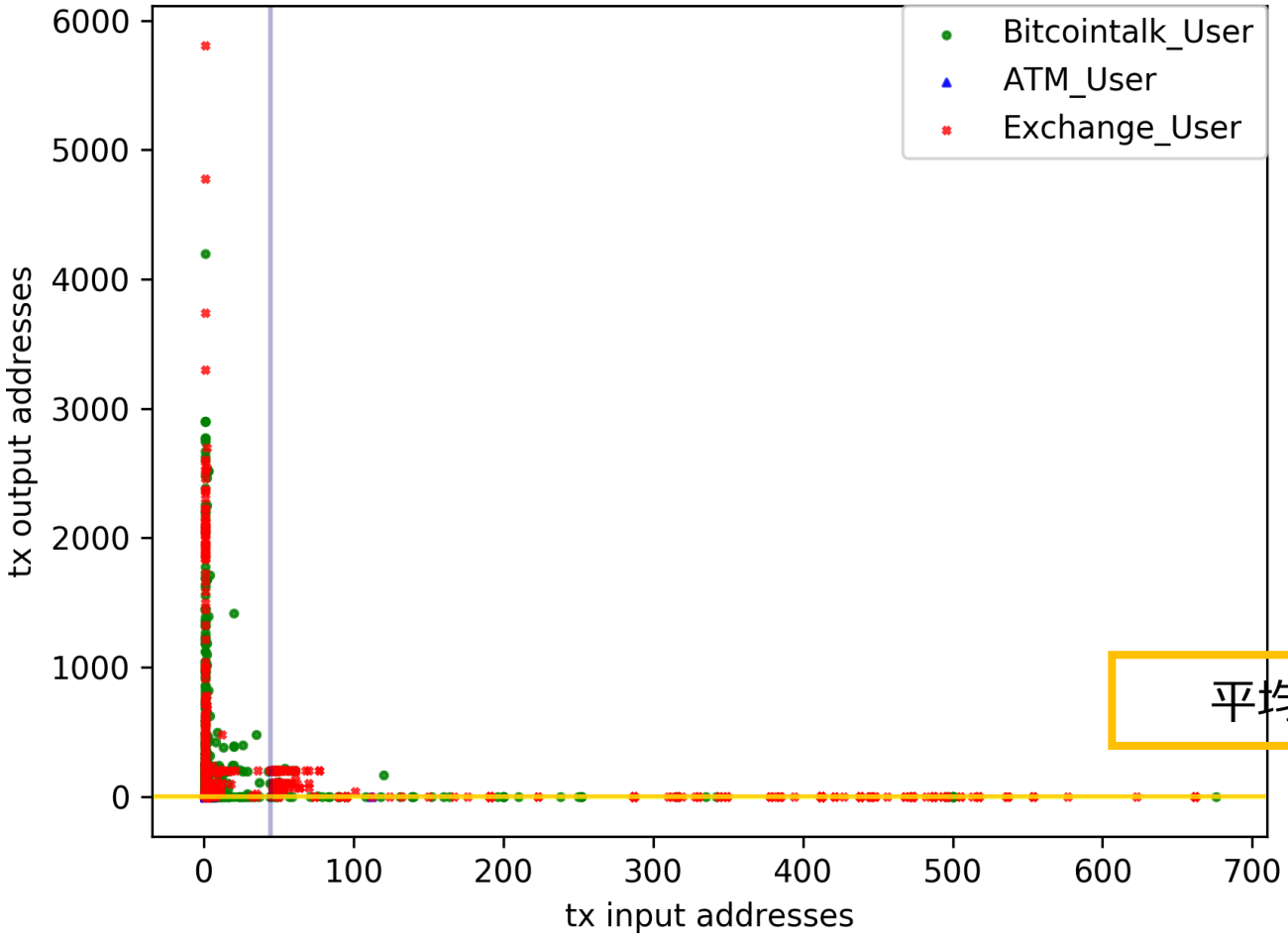
- 以下の条件で枝刈りを実施
  - » ルートからの深さ5以下
  - » 1つの葉に属するサンプルの割合が10%以上
- 100回作成した決定木モデルの一例

3種類のアドレス推定

- 枝刈りを行うと7種類のアドレスを推定することができなかった
  - » **Bitcointalkユーザ**, **ATMユーザ**, **交換所ユーザ**の3種類のみ推定可能
- 3種類のアドレスについて推定基準となる特徴量を分析

3種類のアドレス推定

最小入力アドレス数 ≤ **44.5**個



平均出力アドレス数 ≤ **2.1**個

図　決定木モデルの特徴量と推定アドレス分布

まとめ

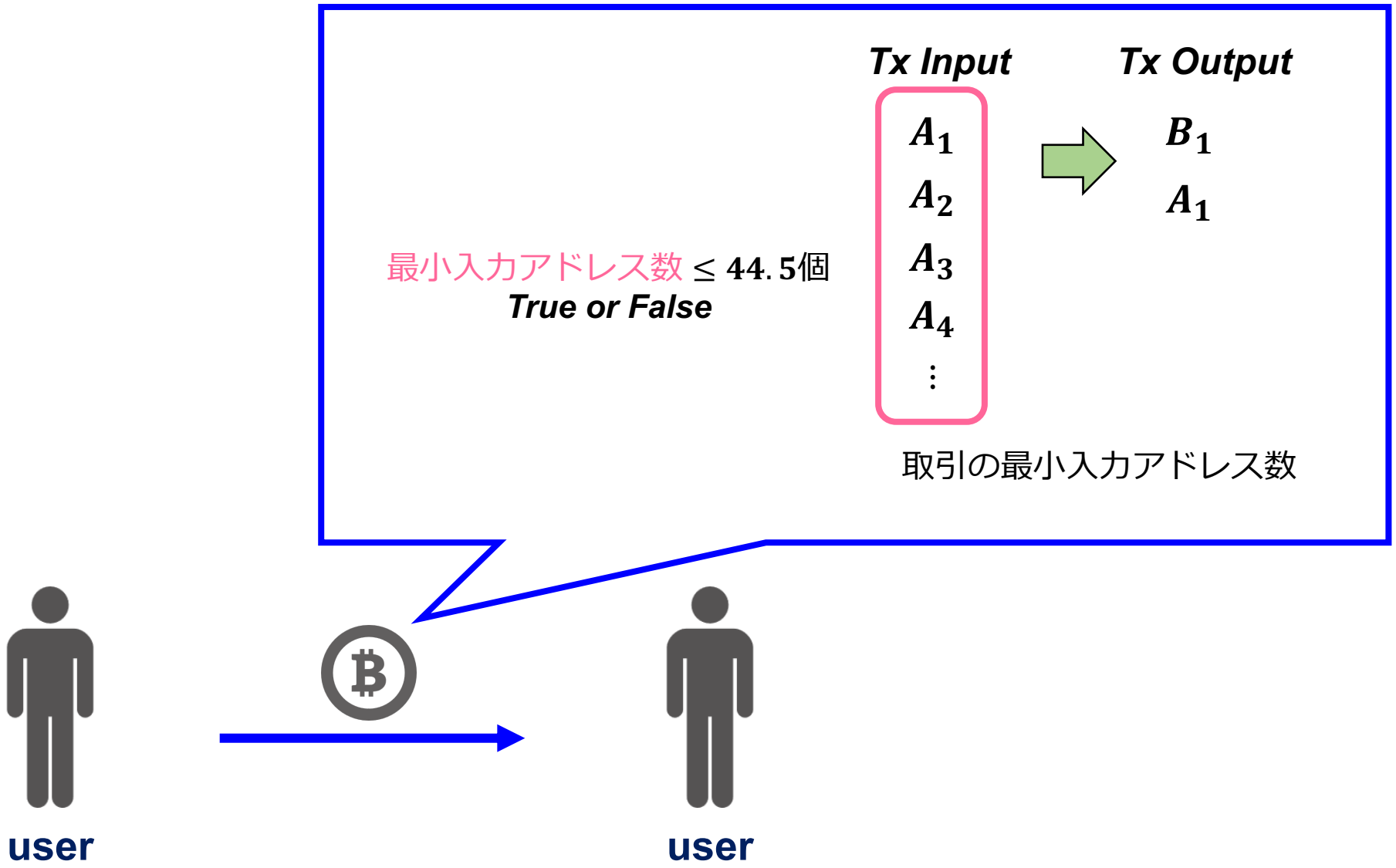
- アドレス種類推定リスクは”**交換所ユーザ**”が最も高い
  - **正解率85%, 適合率80%, 再現率79%**

- アドレス種類の推定に最も影響を与えた特徴量は
  ”**1回の取引で利用される入力アドレス数の最小値**”
  - **44.5個以上のアドレスが入力に使用**された場合 ➔ 交換所ユーザアドレスの**約48%**が該当

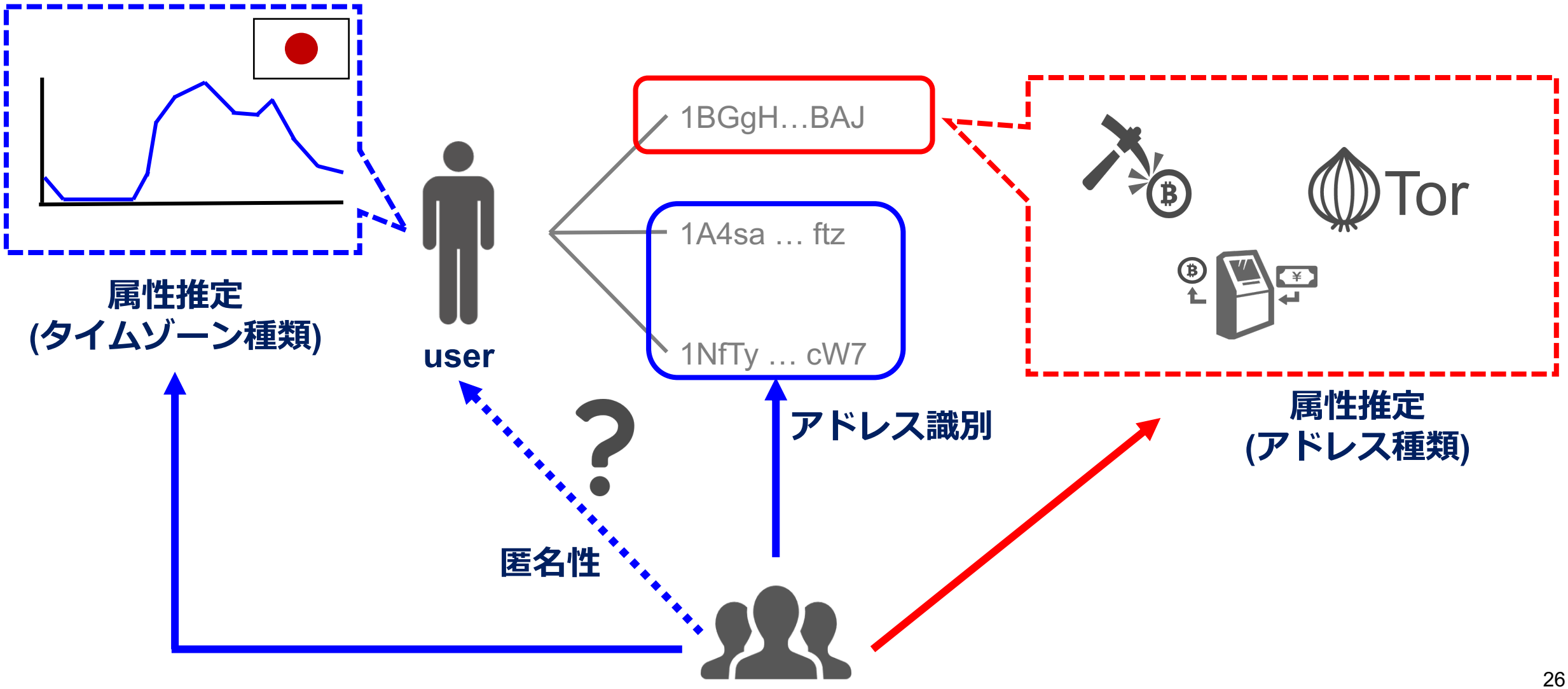
- 課題
  - 7種類のアドレス数にばらつきがある
    - » 最小はBitcoinATM業者の3個, 最多はBitcointalkユーザの2,391個
    - » アドレス数を均一にする, アドレス数に影響しない学習モデルの作成が今後の課題

質疑応答用スライド

ルート特徴量：最小入力アドレス数



アドレスの識別と属性推定



正解率 (*Accuracy*)

- 予測が正しかったもの / 全数

$$Accuracy = \frac{True\ Positive\ +\ True\ Negative}{True\ Positive + False\ Positive + False\ Negative\ +\ True\ Negative}$$

項目名		BitcoinATM 業者	Darkweb 業者	MiningPool 業者	Bitcointalk ユーザ	BitcoinATM ユーザ	Darkweb ユーザ	Exchange ユーザ	合計
		予測							
BitcoinATM 業者	正解	0	0	0	1	0	0	0	1
Darkweb 業者		0	0	0	8	0	0	0	8
MiningPool 業者		0	0	2	19	8	0	0	29
Bitcointalk ユーザ		0	0	0	633	31	0	53	717
BitcoinATM ユーザ		0	0	0	16	119	0	1	136
Darkweb ユーザ		0	0	0	12	3	2	3	20
Exchange ユーザ		0	0	0	56	9	0	239	304