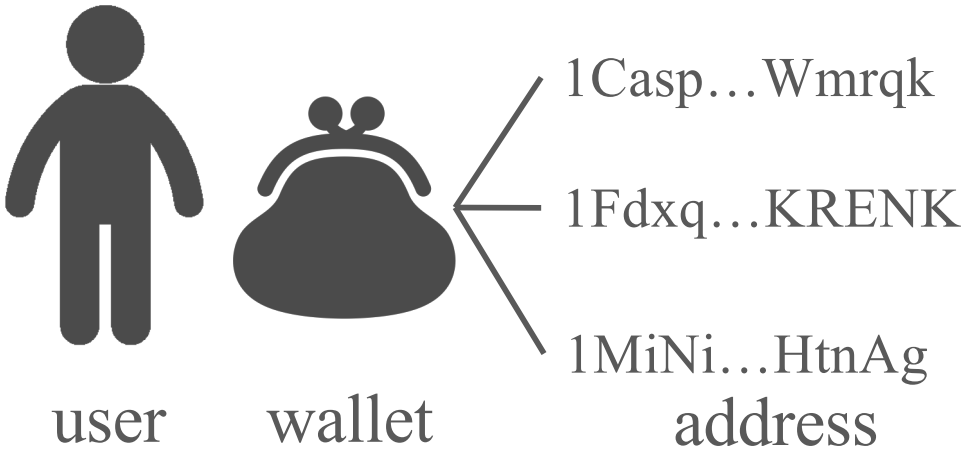
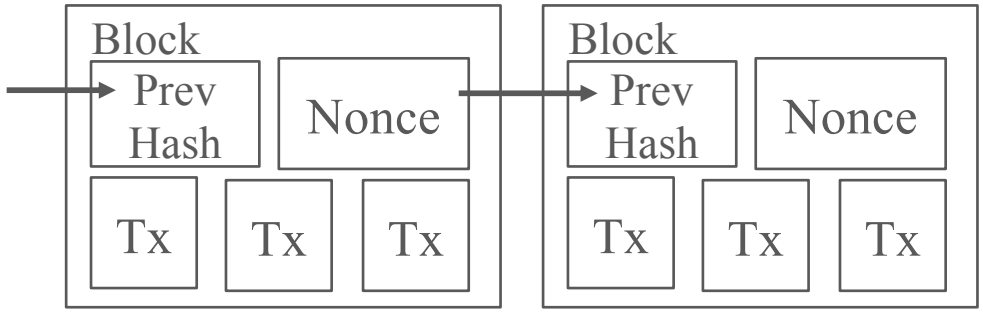
平均取引時間分布の相関を用いた Bitcoin ユーザのタイムゾーン属性の推定

井垣 秀星，永田 倖大，菊池 浩明

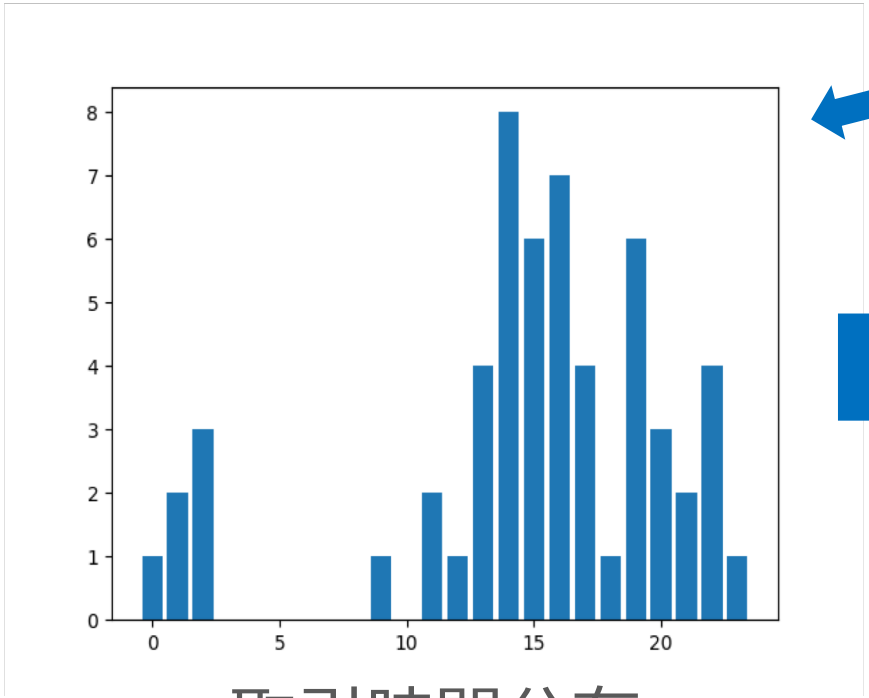
明治大学

本研究の概要





Blockchain



取引時間分布

タイムゾーン属性の推定

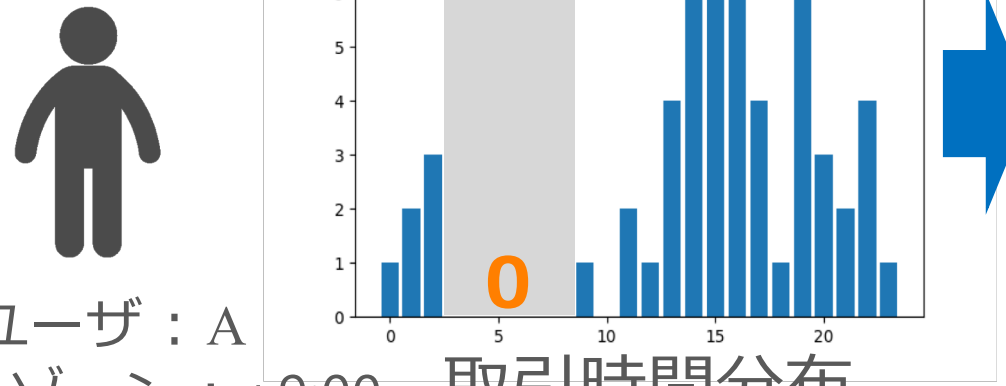
背景:暗号通貨と匿名性

- 2018年1月 コインチェックのNEMが約580億円分不正流出
- 2018年9月 zaifのBitcoin, Monacoin, Bitcoin chacheが合計約70億円分不正流出
  - 誰が盗んだか不明
  - 行方を追うのは難しい

- Bitconの匿名性に関する先行研究
  - 同一ユーザが管理するアドレスを識別(Meiklejohn, IMC'13, 2013)
  - アドレス管理者のタイムゾーンを特定(Dupont, CodaSPY'15, 2015)

先行研究:Bitcoinの匿名性 [Dupont, 2015]

Dupontらのタイムゾーン属性推定の結果は平均推定精度は72%にとどまっていた. 手法は取引が0の時間帯を特徴量とし, タイムゾーン属性の推定を実施.



ユーザ：A
タイムゾーン：+9:00

取引時間分布

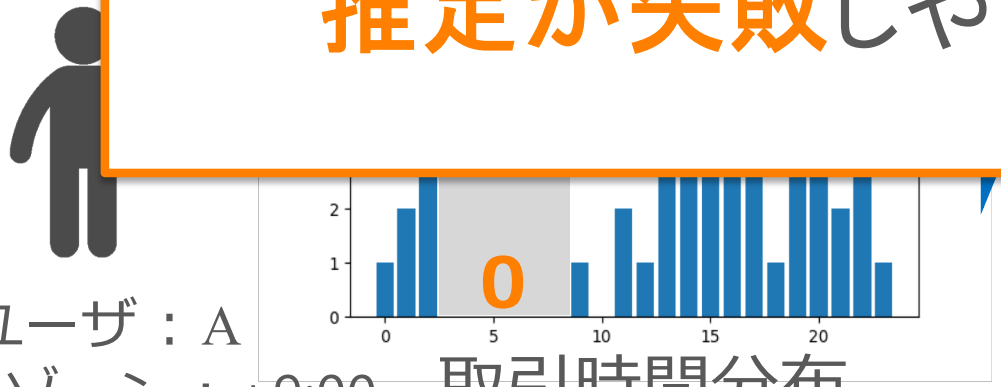
タイムゾーン属性推定

先行研究:Bitcoinの匿名性 [Dupont, 2015]

Dupontらのタイムゾーン属性推定の結果は平均推定精度は72%にとどまっていた. 手法は取引が0の時間帯を特徴量とし, タイムゾーン属性の推定を実施.

小さなノイズにより
推定が失敗しやすい!!

ン属性



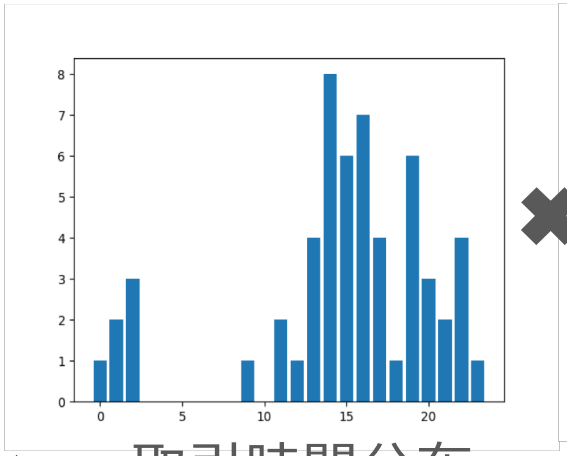
ユーザ：A
タイムゾーン：+9:00

取引時間分布

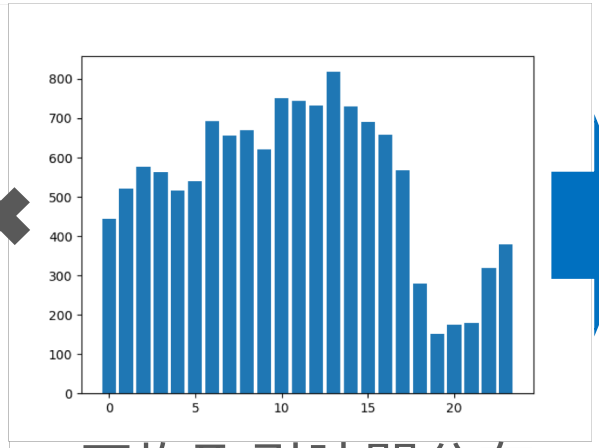
目的:新しいタイムゾーン推定手法の提案

ユーザの取引時間分布を特徴量とし，平均取引時間分布との相関係数に基づく推定方法を提案しBitcoinの匿名性を評価する.

ユーザ：A
タイムゾーン：+9:00



取引時間分布

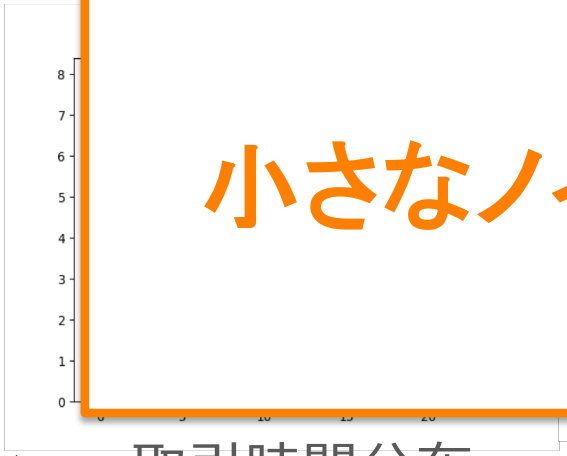


平均取引時間分布

相関係数に基づくタイムゾーン属性推定

目的:新しいタイムゾーン推定手法の提案

ユーザの取引時間分布を特徴量とし, 平均取引時間分布との相関係数に基づく推定方法を提案しBitcoinの匿名性を評価する.



小さなノイズに対して頑強!!

に基づく
ーン属性
定

ユーザ：A　　取引時間分布　　平均取引時間分布

タイムゾーン：+9:00

先行研究と当研究の比較

	当研究	先行研究 [Dupont, 2015]
データ取得期間	**2009年1月3日 - 2018年9月23日**	2009年1月3日 - 2014年9月31日
取得アドレス数	**1, 233**	518
推定手法	**平均取引時間分布との 相関係数を用いた推定手法**	取引数0の部分を 特徴量としたヒューリス ティックな推定手法
精度	**77%**	72%
耐ノイズ性	**高い**	低い

データセットの形式

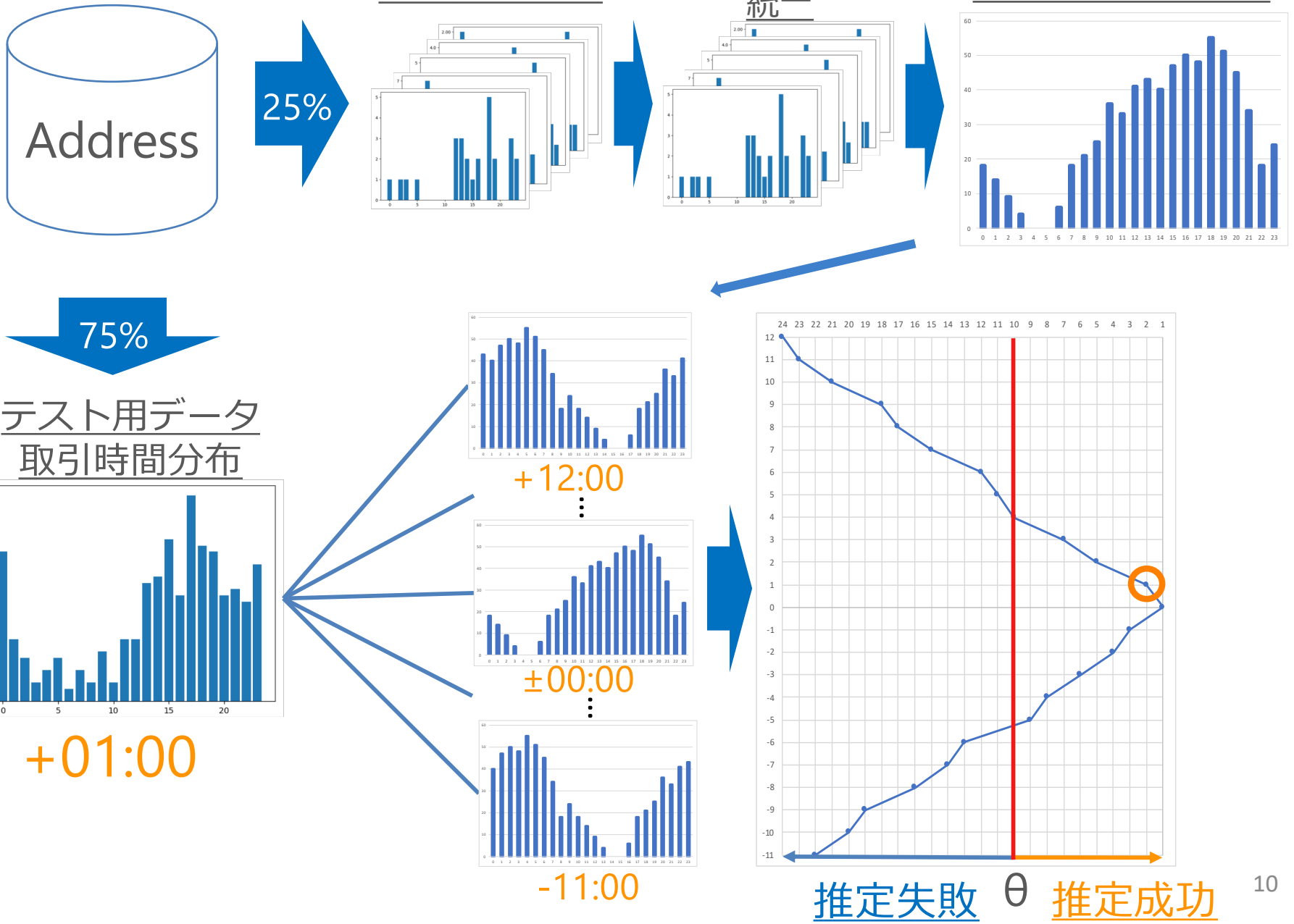
Addressテーブル

Address	Location
1CasperDEhy…	Hong Kong
1FdxQxtzkRR…	France
1MiningaRyy…	Japan

Timestampsテーブル

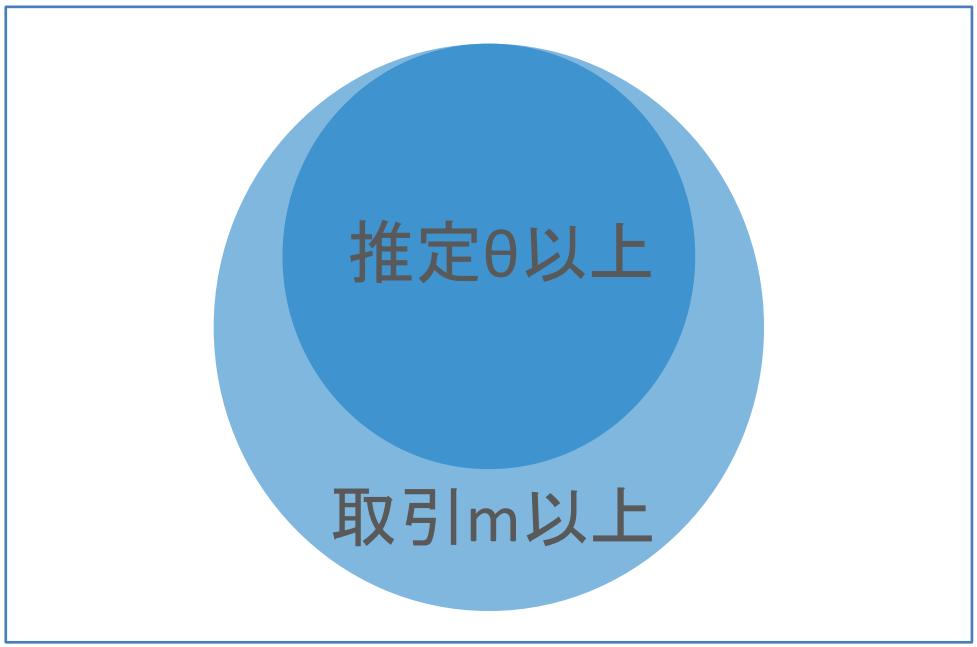
Address	Timestamp
1CasperDEhy…	2012-02-18 00:57:33
1CasperDEhy…	2012-02-18 04:24:46
1FdxQxtzkRR…	2012-06-23 23:45:23
1FdxQxtzkRR…	2013-07-11 13:05:51
1FdxQxtzkRR…	2013-08-09 21:43:21
1MiningaRyy…	2012-04-27 20:32:26
1MiningaRyy…	2012-07-23 01:42:27

提案推定手順


Address
25%
学習用データ
タイムゾーンの統一
平均取引時間分布
75%
テスト用データ取引時間分布
+01:00
+12:00
±00:00
-11:00
推定失敗
θ
推定成功

評価方法：推定成功率

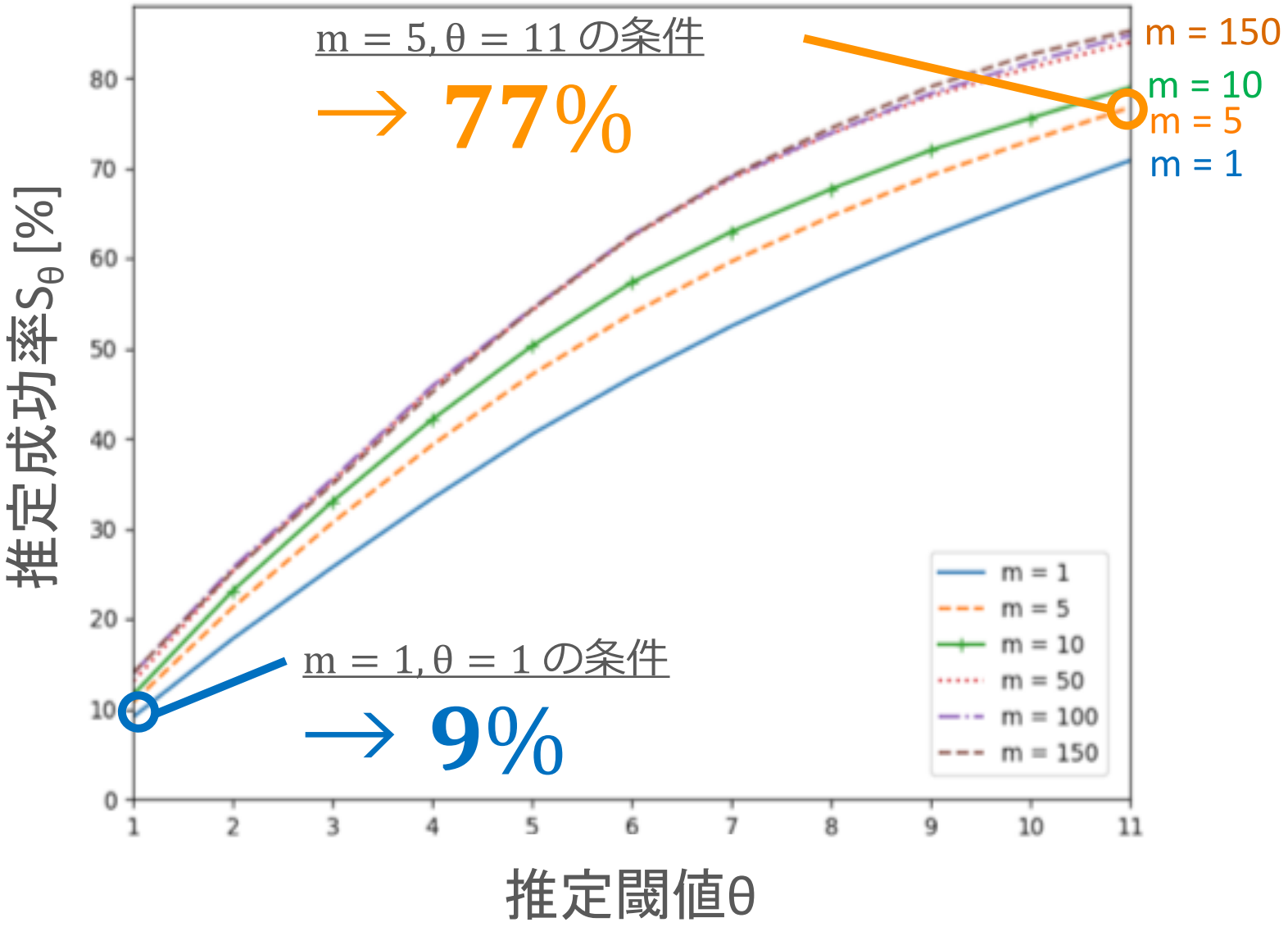
- 推定成功率：s_θ
- Address：i
- 正しいタイムゾーン：i_*
- 推定：j_*
- 取引：t_i
- 取引閾値：m
- 推定閾値：θ



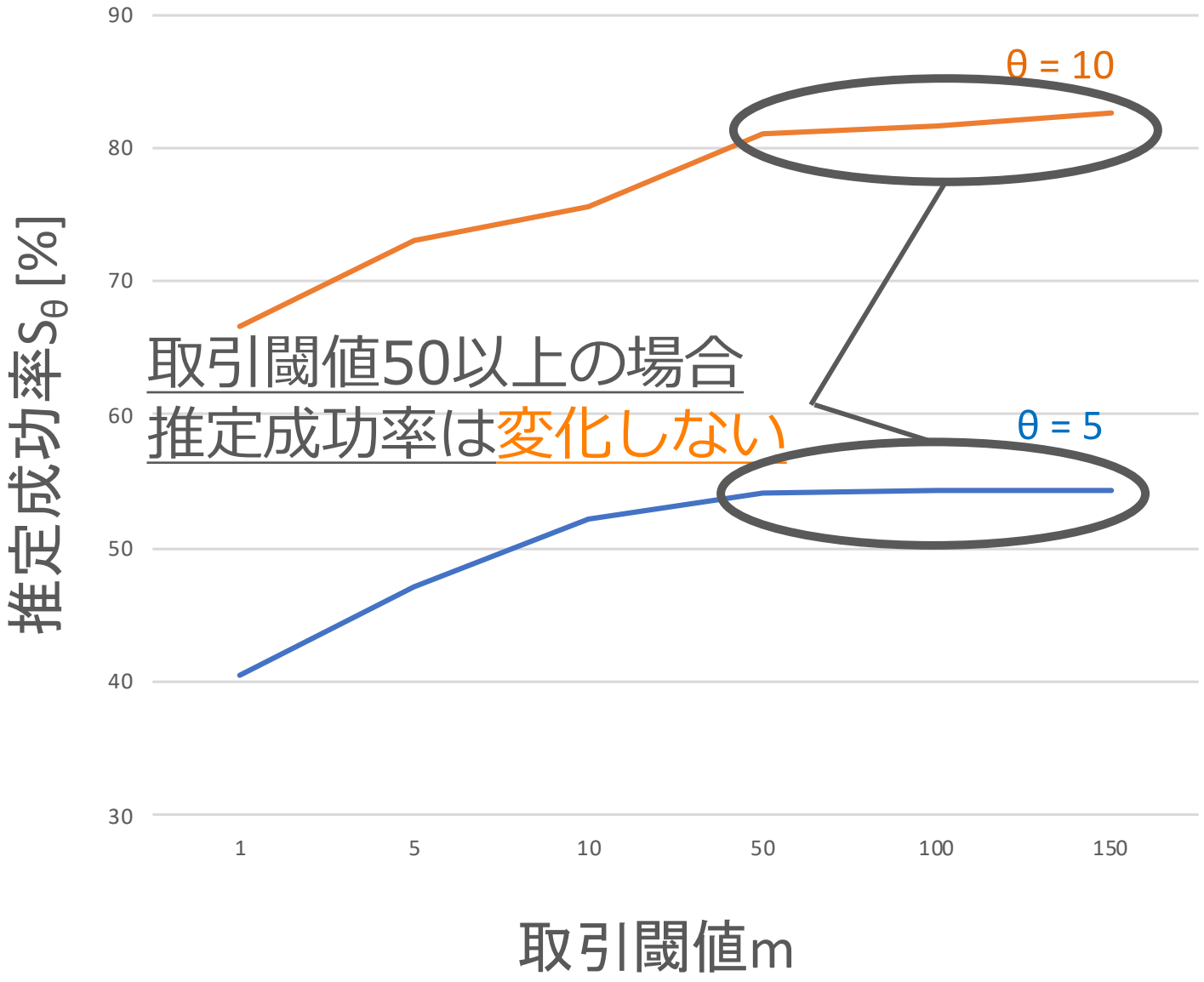
$$s_\theta = \frac{\text{推定}\theta\text{以上かつ取引m以上のユーザ}}{\text{取引m以上のユーザ}}$$

$$= \frac{|\{i \in \mathbb{U} | t_i \geq m, |j_* - i_*| \leq \theta\}|}{|\{i \in \mathbb{U} | t_i \geq m\}|}$$

実験結果1：推定成功率


m = 5, θ = 11 の条件
→ 77%
m = 150
m = 10
m = 5
m = 1
m = 1, θ = 1 の条件
→ 9%
m = 1
m = 5
m = 10
m = 50
m = 100
m = 150
推定成功率S_θ [%]
推定閾値θ
0
10
20
30
40
50
60
70
80
1
2
3
4
5
6
7
8
9
10
11

実験結果2：推定成功率

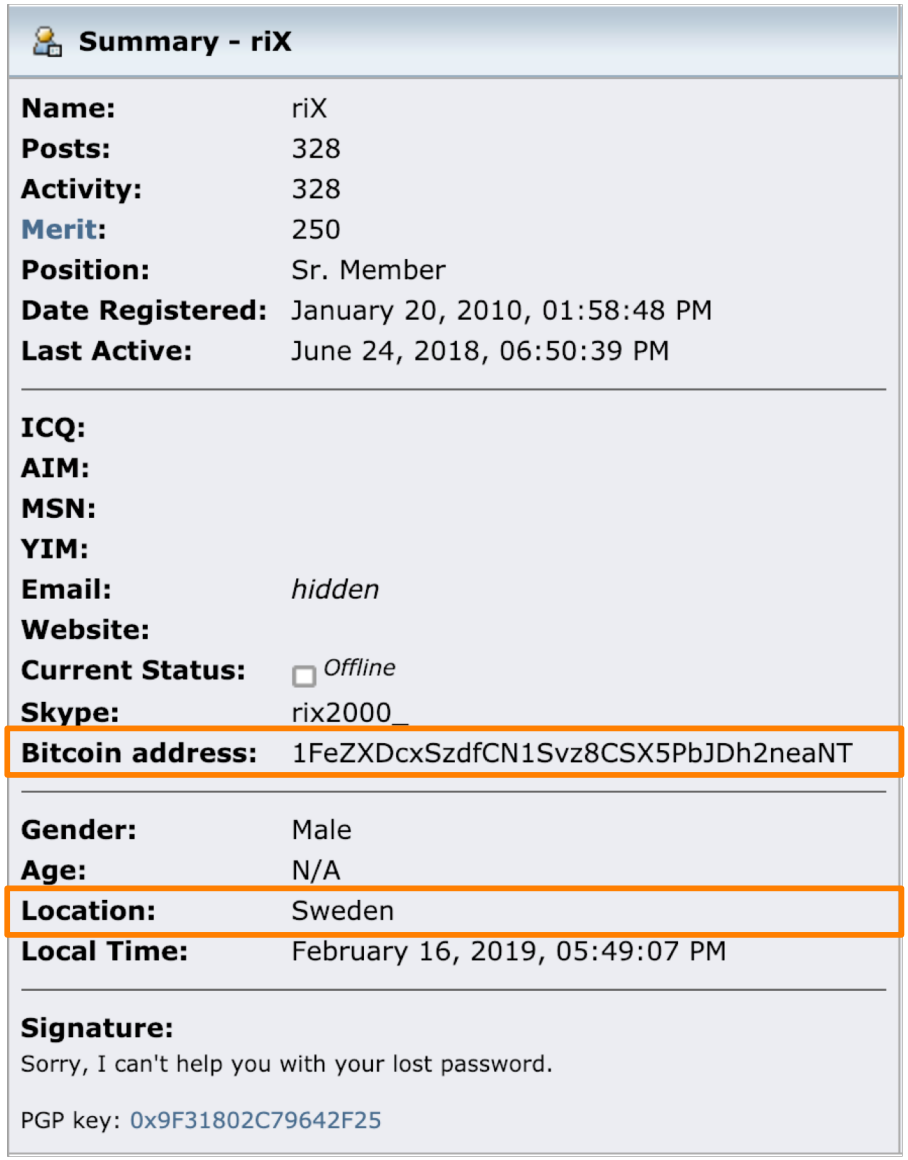

90
80
70
60
50
40
30
推定成功率S_θ [%]
θ = 10
θ = 5
取引閾値50以上の場合
推定成功率は変化しない
1
5
10
50
100
150
取引閾値m

まとめ

- 平均取引時間分布を用いた推定によって，77%のユーザがタイムゾーンを推定されることを示した．
- 取引が50回以上の場合は推定成功率は変わらないことを示した．

質問用スライド1 アドレスについて

サイト名：Bitcointalk



Summary - riX

Name:	riX
Posts:	328
Activity:	328
Merit:	250
Position:	Sr. Member
Date Registered:	January 20, 2010, 01:58:48 PM
Last Active:	June 24, 2018, 06:50:39 PM
ICQ:	
AIM:	
MSN:	
YIM:	
Email:	*hidden*
Website:	
Current Status:	*Offline*
Skype:	rix2000_
Bitcoin address:	1FeZXDcxSzdfCN1Svz8CSX5PbJDh2neaNT
Gender:	Male
Age:	N/A
Location:	Sweden
Local Time:	February 16, 2019, 05:49:07 PM

Signature:
Sorry, I can't help you with your lost password.

PGP key: 0x9F31802C79642F25

→ Addressデータ

→ Locationデータ

[引用] Bitcointalk, https://bitcointalk.org/

質問用スライド2 取引について

サイト名：Blockchain

Timestampデータ

[引用] Bitcoin, https://www.blockchain.com/

質問用スライド3 タイムゾーンについて

- 推定対象とするタイムゾーンは以下の24個とする.

  (+12:00,..., +1:00, ±00:00, -1:00, ..., -11:00)

- サマータイムに関しては考慮しないものとする.

質問用スライド4 データセットについて

Addressテーブル

Address	Parent Address	Location
1CasperDEhyGD81WNPo9qkaFnWxUSWmrqk	none	Hong Kong
1FdxQxtzkRRcCApy7AFGroUgjesyLKRENK	1CasperDEhyGD81WNPo9qkaFnWxUSWmrqk	Hong Kong
1MiningaRyyYRDFGXzwQanfpwd9N5HtnAg	1CasperDEhyGD81WNPo9qkaFnWxUSWmrqk	Hong Kong
…	…	…

Timestampsテーブル

Address	Timestamp
1CasperDEhyGD81WNPo9qkaFnWxUSWmrqk	2012-02-18 00:57:33
1CasperDEhyGD81WNPo9qkaFnWxUSWmrqk	2012-02-18 04:24:46
1FdxQxtzkRRcCApy7AFGroUgjesyLKRENK	2012-06-23 23:45:23
…	…

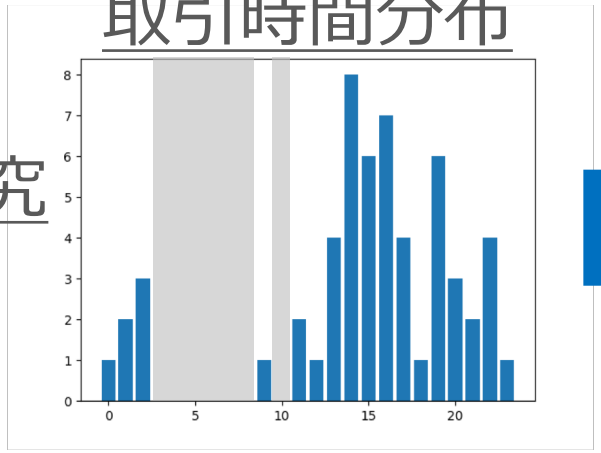
質問用スライド5 耐ノイズ性に関して

例)

タイムゾーン：+4:00

先行研究

取引時間分布



特徴量

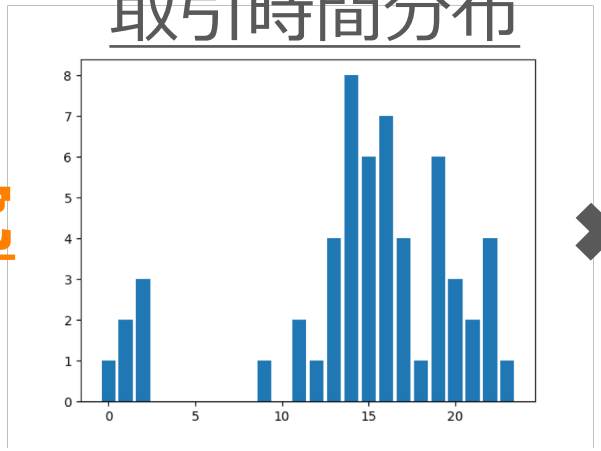
{3, 4, 5, 6, 7, 8, 10}

推定

小さなノイズにより
当ユーザの推定成功率
は下がる

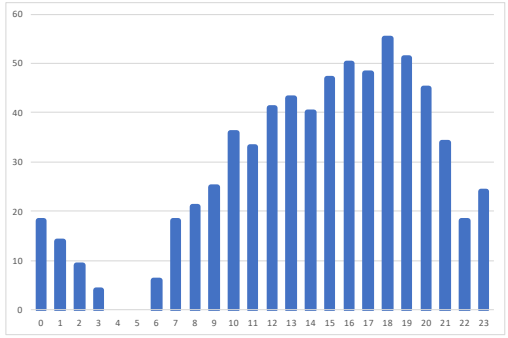
当研究

取引時間分布



✖

平均取引時間分布



推定

小さなノイズでは
当ユーザの推定成功率
に大きな影響はない