

リアル空間でのビットコインの流れ：スキームイベントのケース

全珠美[†] 水野貴之[‡]総合研究大学院大学[†] 国立情報学研究所/総合研究大学院大学[‡]

1.はじめに

現在、ビットコインは時価総額や取引ボリュームが最も高い代表的な仮想通貨である[1, 2]. 全ての取引履歴は保存されており、通貨を取引する仮想空間でのユーザーの取引行動や通貨の流れを、容易に観測することができる。しかしながら、仮想空間での取引を、実空間のどこでおこなっているかを容易に知ることはできない。

実空間に関して、強い匿名性をビットコインは持っているが、ユーザーやビットコインの流れを実空間で追跡することは不可能ではなく、ダークマーケット SilkRoad の事件でも捜査が行われた[3]. 近年、アドレスのクラスタリングからユーザーを特定する手法[4]やビットコインノード間のメッセージからノードの IP を追跡する手法[5, 6]などによる実空間での追跡手法が提案されている。

仮想通貨だけではなく、仮想空間で活動するユーザーの実空間における場所の推定は同じく研究されている。例えば、Twitter における投稿パターンから地域を分類する研究が行われている[7].

本稿では、ビットコインユーザー(アドレス、ウォレット単位)の取引の特徴から、実空間における取引地域を推定し、仮想空間での特定のスキームイベントにおいて、どの実空間の地域のユーザーが被害を受けているのか、または、加担しているのかを推定する。

2.データセット

1つめのデータセットは、仮想通貨コミュニティ Web サイト Bitcointalk.org における 2009 年 12 月から 2018 年 9 月までの、ビットコインに関する 3,251,067 件の投稿記事と投稿日時、891,795 件のユーザー情報である。ユーザー情報にはユーザー名、本 Web サイトの利用頻度、地域(タイムゾーン)が含まれている。本 Web サイト

では地域が GMT に初期設定されているため、GMT のユーザーを解析から除外した。

2つめのデータセットは、www.wallexplorer.com にまとめられているビットコインの取引データである。このデータには、トランザクション ID、直接やり取りしたアドレス、BTC 量、直前・直後のトランザクション ID が含まれている。この Web サイトでは、アドレスをクラスタリングしてウォレット形式で提供している。また、特定のウォレットの場合には、取引種別を表す TAG が付けられている。我々は、仮想通貨コミュニティ Web サイト Bitcointalk.org でアドレスが公開されているユーザーの取引データと、2015 年 4 月から 2016 年 7 月の間に Bitcointalk.org に掲載されたスキームキャンペーン[8]のアドレスと取引したユーザーの取引データに注目する。

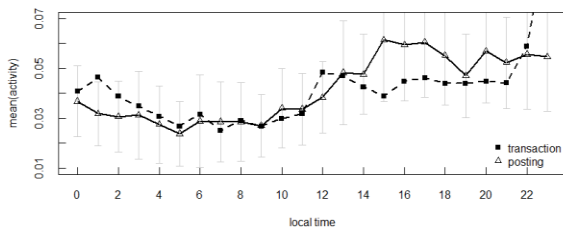
3.ユーザーの実空間における取引地域の推定

人間の活動は 24 時間の中でパターン化している。一日中には必ず休みの時間が発生する。昼休みであったり睡眠であったり、地域(タイムゾーン)の影響を受ける。つまり、活動パターンからユーザーの地域を推定することが可能である。

仮想通貨コミュニティ Web サイト Bitcointalk.org に参加しているビットコイン取引者の、サイトにおける投稿とビットコイン取引の日中の活動パターンに注目する。図 1 は、投稿と取引のパターンを表した例である。パターンが類似していることが読み取れる。

我々は、各ユーザーの地域情報、サイトにおける投稿やビットコイン取引の日中の活動パターンを学習データとして利用し、日中の活動パターンから地域を推定する分類器を構築する。

[図 1]ユーザーのビットコイン取引と投稿の活動パターン



学習データでは、24時間の中で少なくとも5時間以上の時間帯で取引や投稿をおこなったユーザーに限定し、地域をヨーロッパ EU(time zone +1 ~ +3)とアジア ASIA(time zone +7 ~ +9), それ以外の地域 OTH に分類した. ヨーロッパとアジアのユーザーは多く, 全体の 89% (8,600 ユーザー) を占めている. このようにサンプルサイズに偏りがあるため, アップサンプリングをおこない, 学習データの偏りをなくした.

本稿では, 日中の活動パターンから地域を推定する分類器の構築に, 機械学習の一種である XGBoost[9] を採用する. XGBoost は Random Forest にアンサンブル学習である Boosting を結合した手法であり, 現在, 処理速度が優れ信頼性も高い機械学習の1つとして知られている.

表 1 は, 学習データをトレーニングデータセット (80%) とテストデータセット (20%) に別け, XGBoost による学習をおこなった結果である. 90%近い精度で, ユーザーの地域を推計できていることが分かる.

[表 1] time zone 分類学習結果

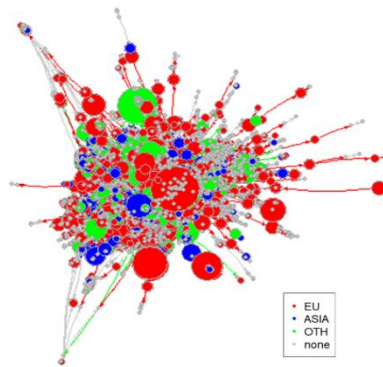
	Post data	Transaction data
Accuracy	0.888	0.91
95% CI	(0.86, 0.912)	(0.871, 0.939)
No information rate	0.333	0.333
Kappa	0.832	0.865

4.スキームイベントと取引ユーザーの地域

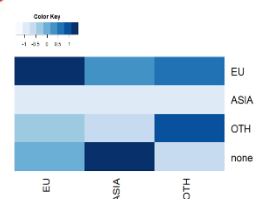
仮想通貨コミュニティ Web サイト Bitcointalk.org に投稿されたスキームイベントにおける, イベント主催者のビットコインアドレスを通過したビットコインの流れを, 図 2 で示す. 流れは, イベント主催者のビットコインアドレスを中心に 6 取引先まで示している. ノードの色はアドレスの地域情報 (EU/ASIA/OTH), エッジの色はビットコインを送ったアドレスの地域の色, ノードサイズは取引回数を表している. このスキームイベントでは, EU からのビットコインの流れ込が多かったことが読み取れる.

図 3 は, このイベントの周辺における地域間のビットコインの流れを表している. このイベントの周辺では, EU 地域や ASIA 地域, その他の地

域 OTH から多くのビットコインが EU 地域に流れ込んでいることが分かる. つまり, このスキームイベントは EU 地域にいるアカウントが EU 地域にいるユーザーをターゲットにして開催しており, EU 地域を中心に世界中から, このアカウントに向けてビットコインが流れ込んでいる. そして, 流れ込んだビットコインは, EU 地域の他のユーザーに流出している.



[図 2] Scam Event Network



[図 3] Transaction Ratio

5.今後の課題

本研究では, 仮想通貨コミュニティ Web サイトに参加するビットコインユーザーの地域情報と, 投稿や取引の活動パターンとの関係を機械学習し, 地域を推定する分類機を作成した. この分類機を用いて, 特定のスキームイベントにおけるビットコインの実空間における流れを明らかにした.

本稿では, 地域の分類をヨーロッパとアジア, その他の地域と 3 分類とし, スキームイベントも限定した. 今後は, 地域分類を増やし, 取引規制の緩い仮想通貨市場における各種イベントでの仮想通貨の流れを把握し, どの実空間の地域のユーザーが被害を受けているのか, または, 加担しているのかを推定していき, 仮想通貨取引における法整備を科学的に支援していく.

参考文献

- [1] Nakamoto, S. (2008) Bitcoin: A Peer-to-Peer Electronic Cash System
- [2] <https://coinmarketcap.com/coins/> (Accessed 2019 Jan)
- [3] John B. (2016) Why criminals can't hide behind Bitcoin. Science Magazine
- [4] Reid, F.; Harrigan, M. (2012) An Analysis of Anonymity in the Bitcoin system. Security and Privacy in Social Networks 197-223
- [5] Kaminsky, D. (2011) Black Ops of TCP/IP, Presentation, Black Hat & Chaos Communication Camp
- [6] Peter L. Juhasz, Jozsef Steger, Daniel Kondor, Gabor Vattay. (2016) A Bayesian Approach to identify Bitcoin Users. PLoS ONE
- [7] Jalal Mahmud, Jeffrey Nichols, Clemens Drews (2014) Home Location Identification of Twitter Users. ACM Transactions on Intelligent Systems and Technology.
- [8] <https://bitcointalk.org/index.php?topic=1021399.0> (Accessed 2019 Jan)
- [9] Tianqi Chen, Carlos Guestrin. (2016) XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)