

機械学習を用いた地域間の仮想通貨フローの可視化

Visualization of Inter-Regional Flows in the Virtual Currency using Machine Learning

全 珠美*¹
Joomi Jun

水野 貴之*^{1,*2}
Takayuki Mizuno

*¹ 総合研究大学院大学
SOKENDAI

*² 国立情報学研究所
National Institute of informatics

Human activity creates a specific pattern in 24 hours. The patterns can be influenced by the time zone they live in. Therefore, we can classify their time zone by their activity pattern. We have built a bitcoin-time zone classifier using XGBoost, a machine learning approach. We have classified the time zone of the specific bitcoin addresses used in the Ponzi event and visualized the flow.

1. はじめに

激しい価値当落やその財貨としての信頼性に疑問を挙げるにも関わらず、仮想通貨への人々の興味や投資今でも続いている。ビットコインの場合、2017年の最高値以来、値段も取引量も下がっているが、いまだ仮想通貨市場で最大の出来高を維持している[CM 19]。我々は代表的な仮想通貨であるビットコインにおける各ユーザーの取引の日中パターンを機械学習により分類することで、ユーザーの活動地域を推定し、サイバー空間で行われている仮想通貨の流れを実空間に可視化する。

ビットコインは実空間に関して強い匿名性を持っていると思われるが、ユーザーやビットコインの流れを追跡することは不可能ではない。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[Reid 12]、ビットコインのノード間のメッセージからノードのIPを推定する手法[Kaminsky 11, Juhasz 16]などが提案されている。このようにビットコインの匿名性を解消しユーザーの空間情報を明らかにする研究は行われている。

サイバー空間でのユーザーの空間情報を推定する手法はソーシャルメディアでもよく研究されている。例えば、Twitterにおけるユーザーの投稿パターンから地域情報を推定することができる[Mahmud 14]。

本稿では、はじめに、仮想通貨コミュニティウェブサイト Bitcointalk.org で活動するビットコインユーザーの活動関連データ(活動地域、投稿履歴、取引履歴)を分類学習機の学習データとして活用し、取引の地域を分類する分類器を作成する。作成した分類器を用いて、活動地域が不明な十分に取引数があるユーザーに関して、取引パターンから実空間における活動地域を推定する。

次に、推定された活動地域を用いて、特定のビットコインの実空間における流れを分析する。我々は2015年から2016年に掛けて Bitcointalk.org に掲載され行われた特定の Ponzi イベントに注目した。ビットコイン市場の Ponzi イベントは、ユーザーがあるアドレスにビットコインを送ると、イベント主催者は集まったビットコインで投資や事業をおこない、ビットコインを送ったユーザーには何パーセントかの利益が上乘せられて返還される。一般的に、イベント参加者が増えると、初期参加者が利益を得る構造があり、乗り遅れまいと多くのユーザーが殺到する。そして、しばしば、後期参加者には、ビットコインの返還がなされない。我々は、被害者のいる Ponzi イベントにおける実空間でのビッ

トコインの流れを可視化し、どの地域で活動するユーザーが利益を、また、被害を受けているのか明らかにする。

2. データセット

ユーザーの活動パターンを分析するため二つのデータセットを用意した。一つ目はビットコインと関連して地域情報が分かるユーザーのデータである。我々は仮想通貨コミュニティウェブサイト 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、そして、その間の全891,795人の投稿者に紐づく、ユーザー名、ユーザーレベル(投稿頻度等)、ユーザーの地域情報(タイムゾーン)を収集した。本サイトの地域情報はGMTに初期設定されているため、地域情報がGMTになっているユーザーは解析から除外した。

二つ目はビットコインの取引履歴データである。公開されているビットコイン取引の情報の中から、ビットコイントランザクションID、直接やり取りしたビットコインアドレス、ビットコインの量、直前・直後のトランザクションID、そして、Bitcointalk.orgでユーザーが公開したビットコインアドレスの取引データを、www.walletexplorer.comを利用して集めた。ビットコインアドレスは、ウォレット形式にクラスタリングされている。また、主要なウォレットは、「取引所」や「マイニング」など取引種別のTAGが付けられている。

3. ビットコインユーザーの地域推定

人間は一日中必ず休みの時間が発生するが、その時間帯は地域(タイムゾーン)によって変わる。日本人とアメリカ人の日中活動をGMTに変換しパターン化すると、その形が異なる。つまり、ユーザーの活動パターンが地域(タイムゾーン)の影響を受けるため、そのパターンで地域を推定することができる。

我々はまず、Bitcointalk.orgで公開されているユーザーのビットコインアドレスデータから、サイトにおける投稿のパターンとビットコイン取引パターンに注目した。Fig. 1は投稿と取引のパターンを表しており、二つのパターンが類似していることが読み取れる。

次に、Bitcointalk.orgのユーザー投稿パターンを用いてビットコイン取引が行われる地域を推定する分類器を構築する。ユーザーの投稿パターンデータと地域情報を学習データセットとした。パターンを十分に学習させるために、学習データは、日中5時間以上の時間帯で投稿を行ったユーザーに限定した。地域の分類は、ヨーロッパEU(GMT+1~3)、アジアASIA(GMT

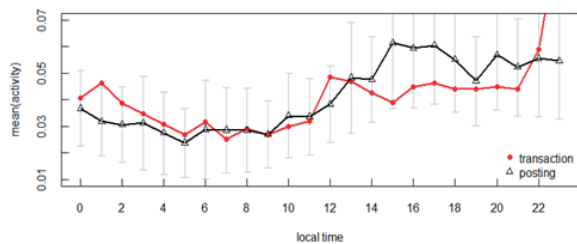


Fig. 1 : ユーザーの投稿パターンとビットコイン取引パターン

	投稿データ	取引データ
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

Table 1 : ユーザー地域分類学習の結果

+7~9)とそれ以外の地域 OTH の三つに分類した。ヨーロッパとアジアのユーザーは今回使ったデータの 89%を占めていた。アップサンプリングをおこなうことにより、サンプルサイズの偏りをなくした。

本稿ではユーザー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[Chen 16]を採用した。XGBoost は Random Forest アルゴリズムを基盤とし、Gradient Boosting を結合した学習法である。速度の遅い Gradient Boosting を並列処理で行っているため学習や分類の信頼性と速度が優れている。我々は、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%) に分け、XGBoost で学習した結果が Table 1 である。地域推定を 90%に近い精度でおこなえていることが分かる。

4. イベントにおけるビットコインの地域間の流れ

構築した地域(タイムゾーン)分類器を利用し、特定の Ponzi イベントにおける地域間のビットコインの流れを分析する。はじめに、イベント主催者のウォレットを中心に、直前(送った)・直後(貰った)取引先の地域を分類する。Fig. 2 の地域別取引回数から、EU ユーザーを対象としたイベントであることが読み取れる。

次に、イベント主催者のウォレットに送る前の3取引先と、主催者のウォレットから受け取った後の3取引先を調査し、地域間の流れを観測する。Fig. 3 は、前・後三つまでの取引先の地域を分類し、各地域間におけるビットコインの出来高(BTC 量)を表している。同地域間の取引が多いことが分かる。つまり、匿名性を持つビットコイン市場であっても、物理的距離の近いユーザー同士が主に取引している。

物理的距離が近いユーザー間の取引が多い理由の一つは、ビットコイン取引ネットワークにおけるハブの存在である。取引所やサービス系会社(ATMやPayment)のウォレットがハブの役割をしており、このハブの活動パターンは、しばしば地域的特徴を強く持っている。各地域のユーザーは、その地域のこのようなハブを好んで利用する傾向があり、その結果が、地域内の取引が多い要因になっている。今回、分析を行った Ponzi イベントの前後の取引で利用された TAG 付がなされたウォレットの中でも、取引所が 24.9%、サービス系会社が 25.3%を占めていた。

5. おわりに

本研究では機械学習手法 XGBoost を利用し、ビットコインコミュニティにおけるユーザーの投稿パターンとビットコイン取引のパターンからユーザーの地域を推定する分類器を作成した。こ

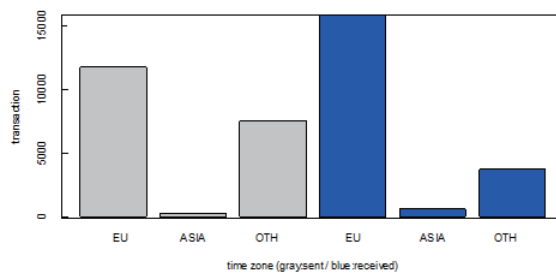
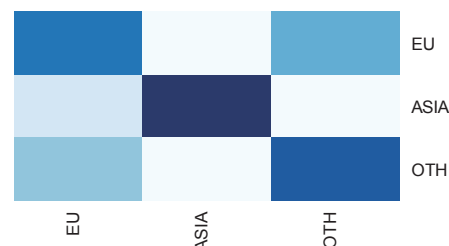
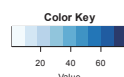


Fig. 2 : イベント参加者の地域別取引回数

Fig. 3 : 地域別ビットコインの流れ(%)
(X : Target / Y : Source)

の分類器を用いて特定の Ponzi イベントに参加したビットコインユーザーの地域を推定し、ビットコインの実空間での流れを調査した。このイベントは、EU ユーザー向けに開催されており、各地域内で主に取引されていたビットコインが、ユーザーがこのイベントに参加することで、EU に流れ込み、その後、それらのコインは EU 内で流通していることが分かった。つまり、このイベントで各国に被害者がいるとすると、被害者のビットコインは主に EU 域内で流通していると言える。

ビットコイン以来、多種多様な仮想通貨が生み出され、その利用範囲は拡大している。仮想通貨は、主にサイバー空間で取引されるが、その取引は実空間での社会生活や経済活動と密接に関係している。この関係性を理解するためには、実空間に射影した流れや利用の特徴を把握する必要がある。本研究のような、仮想通貨の流れを実空間に可視化する努力がこれからも必要になる。

参考文献

- [CM 19] <https://coinmarketcap.com/coins/>(Accessed 2019 Feb)
- [Reid 12] Reid, F.; Harrigan, M. : An Analysis of Anonymity in the Bitcoin system, Security and Privacy in Social Networks, 197-223, 2012
- [Kaminsky 11] Kaminsky D. : Black Ops of TCP/IP, Presentation, Black Hat & Chaos Communication Camp, 2011
- [Juhasz 16] Peter L. Juhasz, Jozsef Steger, Daniel Kondor, Gabor Vattay. : A Bayesian Approach to identify Bitcoin Users. PLoS ONE, 2016
- [Mahmud 14] Jalal Mahmud, Jefferey Nichols, Clemens Drews : Home Location Identification of Twitter Users, ACM Transactions on Intelligent Systems and Technology, 2014
- [Chen 16] Tianqi Chen, Carlos Guestrin. : XGBoost: A Scalable Tree Boosting System, In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016