

الگوی تهیه مقالات برای کنفرانس ملی سالانه محاسبات توزیعی و پردازش داده های بزرگ

صادق دوری نوگورانی^۱، رضا حمیدپور بدوئی^۲

^۱ استادیار، دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس،
Dorri@modares.ac.ir

^۲ دانشجوی ارشد، مهندسی کامپیوتر نرم افزار، دانشگاه تربیت مدرس،
Reza.hamidpour@modares.ac.ir

چکیده

در این مقاله،

کلمات کلیدی

رمزارزهای دیجیتال، مجموعه داده حجیم، مجموعه داده صرافی توزیع شده، مجموعه داده رمزارزهای دیجیتال.

که از جمله مهم ترین این اطلاعات می توان به اطلاعات تبادلات مالی در یک بازار سرمایه^۲ اشاره کرد.

۱- مقدمه

با پیشرفت روز افزون رمزارزها^۱ و افزایش کارایی آن ها در تبادلات مالی، مباحثی همچون تحلیل ساختاری شبکه های زنجیره بلوکی^۲، ایجاد ثبات در تغییر نرخ تبدیل^۳ و همچنین شناخت رفتار کاربران^۴ این شبکه ها بیش از پیش دارای اهمیت شده است. از طرفی نسل جدید تبادلات مالی که بر پایه تبادلات دیجیتالی^۵ شکل میگیرند نیز با پیدایش رمزارزها انقلابی را در نقل و انتقالات مالی ایجاد کرده اند که باعث به وجود آمدن موج جدیدی از تغییرات در بازارهای اقتصادی و چشم اندازهای اقتصادی شده است. با افزایش نیازمندی های امروزه دانشمندان و تحلیل گران اقتصادی به تحلیل دقیق و بهتر ساختار بازار رمزارزها، نیاز به جمع آوری داده این بازارها نیز چالشی بسیار بزرگ است که دانشمندان حوزه تحلیل داده^۶ با آن رو به رو هستند. یکی از این محدودیت ها بدست آوردن مجموعه داده^۷ دقیق از نقل و انتقالات رمزارزها می باشد.

برای توصیف میزان اهمیت داده های اقتصادی می توان به این موضوعات نیز اشاره کرد که با استفاده از این داده ها تحلیل گران اقتصادی، سببگردانان بازارهای سرمایه و همچنین دولت مردان می توانند روند تغییرات یک بازار اقتصادی، میزان تاثیر گذاری قوانین بر روی آن بازار و همچنین روند رشد اقتصادی یک بازار و در امتداد آن یک جامعه را مورد بررسی و تحلیل قرار دهند.

۱-۱ ادبیات موضوعی

در [1] مجموعه داده هایی که توسط رمزارزهای پرطرفدار بیت کوین^۸ و اتریوم^۹ ایجاد شده اند را تحت عنوان مجموعه داده های بزرگ^{۱۰} معرفی میکنند که در حوزه هایی همچون تشخیص تقلب^{۱۱}، تراکنش های غیر قانونی^{۱۲}، شناسایی نقاط بهرانی اقتصاد^{۱۳} و حوزه های پراهمیت دیگری که کاربرد دارند، اشاره می کند. همچنین در این کتاب به توصیف اساسی در خصوص ویژگی های مختلفی که می توان از مجموعه داده های یک رمزارز بدست آورد و یا ساختار کلی این نوع از داده ها پرداخته شده است.

در مقاله [2] نیز به این موضوع اشاره شده است که زنجیره های بلوکی توانایی بسیار زیادی برای استفاده در حوزه مالی، بازی، نقل و انتقالات مالی و زنجیره کالا دارد، اما یکی از چالش های جدی در این حوزه تحلیل رفتاری کاربران این شبکه ها و دسته بندی آن ها بر اساس رفتار شان می باشد. در اینجا نویسندگان با دنباله های تراکنشی کاربران شبکه بیت کوین که از تاریخچه عمومی این شبکه دریافت شده است و با استفاده روش پیشنهادی خود که یک روش تحلیل رفتار شناسی می باشد این داده ها را مورد ارزیابی قرار داده اند.

۱- مجموعه داده

همانگونه که پیش از این نیز اشاره شد با روی کار آمدن رمزارزهای دیجیتال و پر رنگ تر شدن نقش آن ها در دنیای اقتصاد، گرایش تحلیل گران اقتصادی به این نوع از داده ها نیز فزونی یافت. شبکه های بلاکچین با ایجاد تغییر و تحولات گسترده در سیستم انتقالات مالی، بازارهایی را ایجاد کردند که در آن تبادل کنندگان مالی نیازمند یک شخص ثالث که وظیفه تایید و بررسی انتقالات دارایی ها را برعهده داشته باشد ندارند.

پیش از این نیز به این موضوع اشاره شد که در این مقاله مجموعه داده بدست آمده، از مجموعه تبادلات مالی در شبکه استلار می باشد. با توجه به ساختار تراکنش های شبکه استلار داده های پیشنهاد خرید و فروش دارایی به ۲ حالت کلی تقسیم می شوند:

- پیشنهادات خرید یک دارایی
- پیشنهادات فروش یک دارایی

بر اساس توصیف ارائه شده توسط [6] می توان هر پیشنهاد خرید و یا فروش تو سطر یک عامل در بازار را به طور کلی به یک پیشنهاد خرید تبدیل و یا بالعکس. به این صورت که اگر شخصی قصد خرید بیت کوین را دارد و در ازای آن اتریوم ارائه کند، می توان اینگونه تراکنش آن را دید که شخص مورد نظر یک درخواست فروش اتریوم را در شبکه ثبت کرده است.

به همین شکل اگر همان شخص مجددا قصد فروش بیت کوین را داشته باشد درحالی که به ازای آن قصد خرید اتریوم را داشته باشد می توان این بار آن را به عنوان فروشنده اتریوم در نظر گرفت. بر اساس این توصیف در حالت کلی تمامی درخواست هایی که در شبکه استلار ایجاد می شوند، در قالب یک درخواست فروش دیده می شوند.

۱-۱- توصیف فیلدهای داده:

در این بخش به توصیف فیلدهای داده ای هر پیشنهاد در شبکه استلار می پردازیم.

- Source_account: این ویژگی بیانگر ایدی شخصی است که پیشنهاد را درون شبکه ایجاد کرده است.
- Offer_id: در صورتی که این ویژگی برابر با مقدار صفر باشد به این معنی است که این پیشنهاد به تازگی توسط کاربر ایجاد شده است و درون لیت فروش قرار گرفته است.
- Created_at: بیانگر زمان ایجاد این پیشنهاد می باشد.
- Ledger_id: بیانگر ایدی دفتر حساب کلی می باشد که این پیشنهاد در آن قرار گرفته است.
- Transaction_hash: همانگونه که اشاره شد هر پیشنهاد در شبکه استلار در یک تراکنش قرار میگیرد که هر تراکنش مجموعی

در مقاله [3] نیز با جمع آوری مجموعه داده ای از بازار آتی چین^{۱۷} و با استفاده از روش های خوشه بندی اقدام به تحلیل رفتار کاربران این بازار کرده اند.

در این مجموعه داده با توجه به ذات این بازار در هر طرف مبادلات مالی تنها یک دارایی قرار میگیرد که می توان این مورد را یکی از مواردی دانست که در این مقاله به آن توجهی نشده است. چرا که در نسل جدید بازارهای اقتصادی که به وسیله ساختار شبکه های زنجیره بلوکی شکل گرفته اند در تبادلات مالی بین کاربران مختلف دارایی های متفاوتی تبادل می شود.

به عنوان مثال می توان گفت در مجموعه داده جمع آوری شده از شبکه استلار در یک تراکنش ممکن است بین دو کاربر دو رمزارز اتریوم و بیت کوین تبادل شود، اما این درحالی است که در همان لحظه بین دو کاربر دیگر امکان تبادل دو رمزارز دیگر مانند ریپل^{۱۸} و تتر^{۱۹} وجود دارد.

در [4] نیز با جمع آوری داده های پراهمیت رمزارزهای مهمی همچون بیت کوین، اتریوم، لایت کوین^{۲۰} و ریپل که بیشترین میزان سرمایه را در بازار رمزارزها به خود اختصاص داده اند، را مورد تحلیل و بررسی قرار داده اند. در واقع در این مقاله داده های مربوط به تبادلات رمزارزهای مختلف را از بازار (binance.com) جمع آوری کرده اند و سعی در شناخت تاثیرات قوانین اعمال شده بر روی قیمت رمزارزها داشته اند که مهم ترین این قوانین، قوانین اعمال شده در اواخر سال ۲۰۱۷ و اوایل سال ۲۰۱۸ می باشد. در این تحقیق با شناخت الگوی تراکنش های این رمزارزهای مهم، سعی در شناخت هرچه بهتر این بازارها دارند.

مجموعه داده ای که در این مقاله ما جمع آوری کرده ایم، شامل تراکنش ها و مبادلات مالی بین کاربران شبکه استلار^{۱۴} می باشد. این شبکه با بهره گیری از فناوری بلاکچین نسل جدیدی از صرافی های توزیع شده را ایجاد می کند به گونه ای که کاربران این شبکه بدون نیاز به یک شخص ثالث اقدام به تبادل مالی با یکدیگر می کنند.

عملیات ها در شبکه استلار: عملیات ها در شبکه استلار در قالب تراکنش هایی جمع می شوند به گونه ای که هر تراکنش شامل چندین عملیات می تواند باشد (عملیاتی که همچون تغییر یک حساب کاربری، ثبت پیشنهاد خرید/فروش یک دارایی، ویرایش/حذف یک پیشنهاد و انتقال مالی بین کاربران).

دفتر حساب های کل: دفتر حساب های کل^{۱۶} نقش ذخیره سازی تراکنش ها در شبکه استلار را دارند، به گونه ای که در هر دفتر حساب کل چندین تراکنش می تواند قرار بگیرد، به طوری که هر تراکنش نیز شامل چندین عملیات می باشد.

جدول-۱ تعداد تراکنش ها بر اساس دارایی هایی که در قسمت فروش قرار دارند.

تعداد تراکنش	نوع دارایی	تعداد تراکنش	نوع دارایی
26747	MOBI	365124	native
21498	EURT	168119	ETH
18569	CNY	156867	XRP
16198	MERA	143540	BTC
15901	FRAS	128948	LTC
14464	XCN	112107	SLT
13089	ZRX	111378	RMT
10034	USDM	110495	NRV
7659	XFF	40854	TERN
7461	USD	31470	GTN
29500	KIN	29500	KIN

جدول-۱ و جدول-۲ نشان دهنده این موضوع هستند که در شبکه استلار بیشترین حجم تبادلات مالی به ترتیب مربوط به دارایی های بومی شبکه استلار، اتریوم و بیت کوین می باشند.

در مقالات مشابه که بر روی داده های اقتصادی بازارهای مشابه فعالیت انجام داده اند ساختار بازارها به گونه ای بوده است که در آن به ازای دارایی های مختلفی که هر کاربر خریداری می کند تنها یک دارایی واحد را به ازای آن ها پرداخت می کند، به عبارتی دیگر در هر تبادل مالی یک سوی آن تبادل یک دارایی مشخص و واحد می باشد، این درحالی است که در بازار تبادلاتی استلار به ازای دارایی های مختلفی که کاربران خریداری می کنند، دارایی های متفاوت دیگری را ارائه می کنند. به عنوان مثال شخصی که قصد خرید بیت کوین را دارد می تواند در مقابل آن دلار آمریکا ارائه کند، درحالی که شخصی دیگر در همان لحظه می تواند پیشنهاد مشابهی را ارائه کند با این تفاوت که دارایی ارائه شده توسط این شخص به جای دلار آمریکا رمزارز ریپل باشد.

جدول-۲ تعداد تراکنش ها بر اساس دارایی هایی که در قسمت خرید قرار دارند

تعداد تراکنش	نوع دارایی	تعداد تراکنش	نوع دارایی
11943	SALT	1172217	native
10368	OMG	45292	NRV
7649	USD	41608	ETH
7620	REP	33679	XRP
6813	XFF	25590	BTC
6254	LTC	25469	TERN
5986	CNY	22436	GTN
5398	MTL	21722	SLT
5238	LINK	19021	MOBI
4761	BCH	14711	USDM
4693	CALL	13007	XCN

۳- تحلیل داده

در این بخش با تمرکز بر روی دو رمزارز بیت کوین و اتریوم قصد داریم توصیف بهتری نسبت به تبادلات مالی این مجموعه داده ارائه بدهیم. در همین جهت از میان عملیات های انجام شده در این مجموعه داده تبادلاتی که، حداقل در یک طرف آن ها یکی از رمزارزهای بیت کوین و یا اتریوم قرار دارد را در قالب دو گروه قرار می دهیم.

از چندین عملیات است، در این فیلد مقدار هش^{۲۳} تراکنشی که این عملیات متعلق به آن هست قرار میگیرد.

- **Selling_asset_type**: بیانگر نوع دارایی است که پیشنهاد دهنده قصد فروش آن را دارد. در صورتی که مقدار آن برابر با دارایی بومی^{۲۴} در داخل شبکه استلار باشد دو فیلد بعدی مقداری ندارند.
- **Selling_asset_code**: نماد دارایی که به فروش می رسد.
- **Selling_asset_issuer**: ایدی شخصی است که این دارایی را در شبکه ایجاد کرده است. با توجه به ساختار شبکه استلار هر دارایی را یک کاربر می تواند ایجاد کند و به وسیله واسطه هایی^{۲۵} که در شبکه وجود دارند.
- **Amount**: بیانگر مقداری است که شخص پیشنهاد دهنده از دارایی مورد نظر برای فروش ارائه می کند.
- **Price_r {n,d}**: این ویژگی شامل دو مقدار n و d می باشد و میزان ارزش دارایی است که به فروش می رسد در مقابل دارایی که خریداری می شود. به عنوان مثال اگر شخصی بخواهد مقدار ۱ بیت کوین را در ازای ۱۰۰ اتریوم به فروش برساند مقدار این ویژگی برابر با $\{1,100\}$ می باشد.
- **Price**: این ویژگی از ضرب دو ویژگی n در فیلد $price_r$ در ویژگی $amount$ به وجود می آید.
- **Buying_asset_type**: بیانگر نوع دارایی است که شخص پیشنهاد دهنده قصد دریافت آن را دارد در مقابل دارایی که به فروش می رسد. در صورتی که مقدار آن برابر با دارایی بومی^{۲۴} در داخل شبکه استلار باشد دو فیلد بعدی مقداری ندارند.
- **Selling_asset_code**: نماد دارایی که به فروش می رسد.
- **Buying_asset_code**: کد دارایی که پیشنهاد دهنده می خواهد دریافت کند.
- **Buying_asset_issuer**: ایدی شخصی است که در شبکه این دارایی را ایجاد کرده است و پیشنهاد دهنده از آن استفاده می کند.

داده های مورد استفاده در این مقاله به وسیله ای پی آی^{۲۶} های ارائه شد توسط سرورهای هوریزون^{۲۷} شبکه استلار جمع آوری شده اند. بازه زمانی داده های جمع آوری شده برابر است با ۲۰۱۸/۱۲/۰۱ تا ۲۰۱۹/۱۲/۳۰ که در این بازه زمانی ۱،۶۵۰،۷۱۸ عملیات خرید/فروش ایجاد شده است که از این جمع عملیات ۱،۱۶۵،۲۷۴ عملیات درخواست خرید جدید بوده است و ۸۹۱،۴۴۹ عملیات نیز عملیات ویرایش و یا حذف درخواست های قبلی می باشد. تمامی این عملیات ها در ۶،۶۸۰،۴۵۱ دفتر حساب کل جمع آوری شده است. در کل این مجموعه داده ۱۶،۳۳۸ کاربر عملیات تبادل دارایی انجام داده اند به طوری که ۴۶۶ دارایی در حالت فروش قرار گرفته است و تعداد ۴۴۴ دارایی نیز خریداری شده است.

مراجع

- [1] S. S. Chawathe, "Clustering Blockchain Data," in Clustering Methods for Big Data Analytics: Techniques, Toolboxes and Applications, O. Nasraoui and C.-E. Ben N'Cir, Eds. Cham: Springer International Publishing, 2019, pp. 43–72.
- [2] B. Huang, Z. Liu, J. Chen, A. Liu, Q. Liu, and Q. He, "Behavior pattern clustering in blockchain networks," Multimed Tools Appl, vol. 76, no. 19, pp. 20099–20110, Oct. 2017, doi: 10.1007/s11042-017-4396-4.
- [3] G. Shi, L. Ren, Z. Miao, J. Gao, Y. Che, and J. Lu, "Discovering the Trading Pattern of Financial Market Participants: Comparison of Two Co-Clustering Methods," IEEE Access, vol. 6, pp. 14431–14438, 2018, doi: 10.1109/ACCESS.2018.2801263.
- [4] M. Gidea, D. Goldsmith, Y. Katz, P. Roldan, and Y. Shmalo, "Topological recognition of critical transitions in time series of cryptocurrencies," Physica A: Statistical Mechanics and its Applications, p. 123843, Jan. 2020, doi: 10.1016/j.physa.2019.123843.

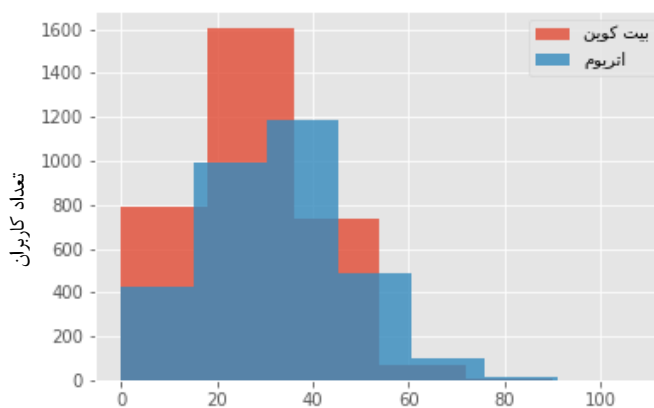
زیر نویس ها

جدول-۳ تعداد تراکنش در هر گروه دسته بندی

نوع دارایی	تعداد پیشنهادات
بیت کوین	۱۶۹.۱۳۰
اتریوم	۲۰۹.۷۲۷

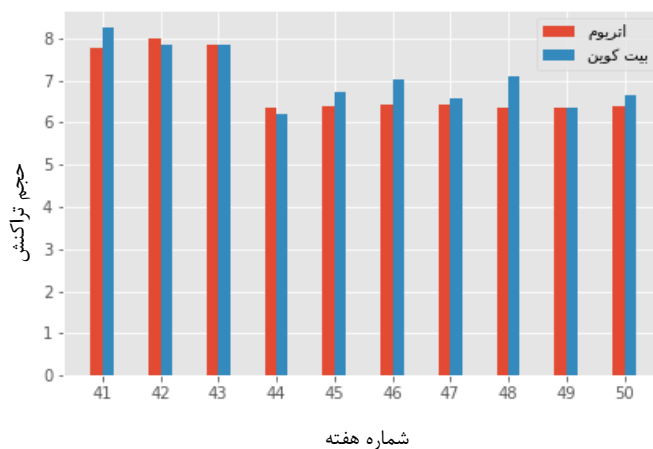
در جدول-۳ تعداد تراکنش ها به تفکیک هر یک از دارایی های بیت کوین و اتریوم قرار دارد که بیانگر تمامی عملیات هایی می باشد که حداقل یک طرف آن ها بیت کوین و یا اتریوم قرار دارد.

شکل-۱ هیستوگرام تعداد تراکنش کاربران بیت کوین و اتریوم



شکل-۱ نشان دهنده تعداد عملیات های انجام شده توسط کاربران شبکه استلار بر روی دو دارایی بیت کوین و اتریوم می باشد. می توان به این موضوع اشاره کرد که میزان پراکندگی تعداد تراکنش توسط کاربران بر روی دارایی اتریوم از توزیع بهتری برخوردار است.

شکل-۲ حجم تراکنش کاربران به تفکیک هر هفته



۴- نتیجه