Blockchain Intelligence: When Blockchain Meets Artificial Intelligence

Zibin Zheng, *Senior Member, IEEE*, Hong-Ning Dai, *Senior Member, IEEE*

Abstract—**Blockchain is gaining extensive attention due to its provision of secure and decentralized resource sharing manner. However, the incumbent blockchain systems also suffer from a number of challenges in operational maintenance, quality assurance of smart contracts and malicious behaviour detection of blockchain data. The recent advances in artificial intelligence bring the opportunities in overcoming the above challenges. The integration of blockchain with artificial intelligence can be beneficial to enhance current blockchain systems. This article presents an introduction of the convergence of blockchain and artificial intelligence (namely blockchain intelligence). This paper also gives a case study to further demonstrate the feasibility of blockchain intelligence and point out the future directions.**

Keywords—Blockchain; Artificial Intelligence; Smart Contract; Machine Learning

I. Introduction

Blockchain has received extensive attention recently due to its provision of secure data sharing services with traceability, immutability and non-repudiation. Despite the merits of blockchain, the development of blockchain technologies has undergone a number of challenges including poor scalability, difficulties in operational maintenance, detecting vulnerable codes in smart contracts and identifying malicious behaviours in blockchain historical data.

The recent advances in artificial intelligence (AI) have greatly propelled the evolution of diverse business applications. The integration of AI with blockchain has the potentials to overcome the limitations of blockchain. We name the intelligent capability bestowed by AI to blockchain as *blockchain intelligence*. In particular, AI (such as machine learning) approaches may help to capture the abnormal behaviours in blockchain after analyzing the blockchain data, detecting and identifying possible vulnerable program codes in smart contracts.

This article aims at reviewing blockchain technologies as well as AI technologies, presenting an in-depth analysis on integrating blockchain with AI, providing implications of enabling technologies of blockchain intelligence. In summary, the main contributions of this article are summarized as follows:

- We first give an overview of blockchain technology and point out the challenges in existing blockchain systems.
- We then review the advances in AI and formally introduce the convergence of AI and blockchain followed by a discussion on opportunities brought by blockchain intelligence.
- We next present a case study to demonstrate the feasibility of blockchain intelligence.

II. Overview of blockchain technologies

As a disruptive software technology, blockchain is reshaping diverse business sectors. Blockchain is essentially a chain-like data structure storing transactions verified by majority of nodes throughout the whole network as shown inf Figure 1. Since the committed transactions in the blockchain have been stored at every node, they are extremely difficulty to be altered or falsified. Integrating with digital signature and asymmetric encryption, blockchain data is authenticated and auditable, implying non-repudiation of the transaction initiator. The length of blockchain keeps growing with the new validated transaction being appended at end of the chain. Data analytics on blockchain data can potentially extract valuable information.

The development of blockchain technologies has experienced two phases: 1) blockchain 1.0 (*i.e.*, symbolized by digital currency) and 2) blockchain 2.0 (*i.e.*, symbolized by smart contracts) [1]. In blockchain 1.0, blockchain has been mainly used for digital currencies like Bitcoin. The appearance of blockchain has promoted the development of smart contracts. A smart contract essentially consists of a number of computerized contractual agreements consented by multiple parties [2]. The contractual clauses embedded in smart contracts will be triggered and automatically executed when a certain condition is satisfied (*e.g.*, who breaches the contract will be automatically imposed with a fine).

Smart contracts have been implemented with the support of blockchain as shown in Figure 1. The approved contractual clauses are converted into executable computer programs. The logical connections between contractual clauses have also been preserved in the form of logical flows in programs (*e.g.*, `if-else-if` statement). The execution of each contract statement is recorded as an immutable transaction stored in the blockchain. Meanwhile, smart contracts guarantee appropriate access control and contract enforcement. In particular, developers can assign access permission for each function in the contract.

Although blockchain and blockchain-enabled smart contracts are promising in reshaping various industrial sectors, the intrinsic limitations of blockchain systems also lead to the following challenges.

1) Operational maintenance. Due to the decentralization and heterogeneity of blockchain systems, it is difficult to identify

Z. Zheng is with School of Data and Computer Science, Sun Yat-sen University, China (email: zhzibin@mail.sysu.edu.cn).

H.-N. Dai is with Faculty of Information Technology, Macau University of Science and Technology, Macau (email: hndai@ieee.org).

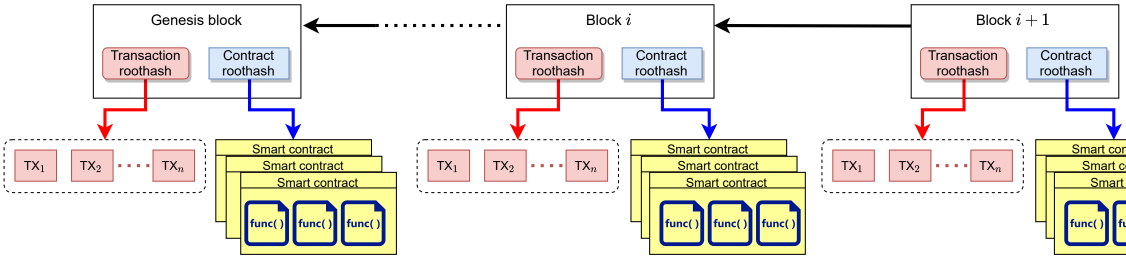


Fig. 1. Overview of blockchain

the potential factors affecting the performance of blockchain. For example, the *transaction throughput* bottleneck of Hyperledger Fabric is different from that of Bitcoin and Ethereum since different consensus algorithms are adopted. Moreover, like other software systems, smart contracts consist of a number of computer programs, which may suffer from software bugs, malicious codes and incompatibility of running environments. Consequently, it is crucial to achieve the intelligent and robust operational maintenance of complex blockchain systems.

2) Quality assurance of smart contracts. Smart contracts suffer from a number of software vulnerabilities such as *reentrancy* vulnerability [1], overcharging issue [3], randomness controlling [4] and Decentralized Autonomous Organization (DAO) attack [5]. In addition, contract correctness is also crucial to smart contracts since it is nearly impossible to make any revisions once they are deployed on top of blockchains. However, like software systems, smart contracts often contain programming bugs which may lead to crashes or misbehaviours while it is challenging to detect and identify these bugs due to the complexity of smart contracts.

3) Malicious behaviour detection. Besides legal businesses, blockchain may be exploited for malicious activities which are nevertheless difficult to be detected due to the *pseudonymity* of blockchain (*i.e.*, anonymous blockchain addresses). On the other hand, the encrypted blockchain data also leads to the difficulty of detecting and identifying malicious behaviours via simply data analytics. Moreover, the massive volume and heterogeneity of blockchain data as well as diversity of user behaviours make the problem even worse. As a result, conventional classification-based methods (*e.g.*, machine learning methods) cannot be directly applied.

III. Opportunities brought by Artificial Intelligence

Artificial intelligence (AI) as a broad discipline covering machine learning and cognitive computing is an ability of intelligent agents conducting intellectual tasks. The recent advances in big data, deep neural networks (DNNs) and general purpose computer hardware such as graphic processing units (GPUs) have greatly driven the development of AI. Consequently, we have witnessed the proliferation of diverse AI applications such as computer vision, natural language processing, speech recognition, sentimental analysis. Big data plays a critical role in propelling AI as well as AI applications. For example, deep learning (mainly based on DNNs) has achieved superior performance thanks to the availability of massive data so that DNNs can extract (or learn) enough features from large volume of data.

The appearance of diverse blockchain systems has generated the enormous volumes of blockchain data, which are publicly available to everyone. Take Bitcoin as an example. As reported by Statista (https://www.statista.com/) at the end of the third quarter of 2019, Bitcoin contains nearly 242 GB data. Data analytics on the massive blockchain data cannot only extract huge business values but also bring great opportunities to overcome the aforementioned challenges of blockchain systems. The recent advances in AI have also greatly driven the development of big data analytics [7]. Thus, the integration of AI and blockchain technologies can potentially overcome the aforementioned challenges of blockchain systems, thereby forming intelligent blockchain systems. We name such integration of blockchain with AI as *blockchain intelligence*.

We summarize the opportunities brought by AI to enhance incumbent blockchain systems as follows (as shown in Figure 2).

1) Intelligent operational maintenance of blockchain. Blockchain generates huge amount of data in a real-time manner. Analyzing blockchain data, we can detect the possible faults, forecast the failures and identify the performance bottleneck so as to tune or adjust the performance of blockchain systems. There are four different levels of data analytics including: descriptive analytics, diagnostic analytics, predictive analytics and prescriptive analytics. In particular, the work of [8] presents a common platform to evaluate the performance of three representative blockchain systems: Ethereum, Parity and Hyperledger Fabric via descriptively analyzing blockchain data. Meanwhile, the descriptive analytics of blockchain log data can help to monitor the real-time performance of blockchain systems and identify the possible faults [6]. In addition to diagnostic analytics on blockchain data, predictive analytics is also necessary to anticipate the performance bottleneck of blockchain systems. Unlike diagnostic and predictive analytics, prescriptive analytics can simulate and optimize blockchain systems so as to improve the reliability of blockchain systems.

2) Intelligent quality assurance of smart contracts. Like computer software, smart contracts may contain bugs or faulty programming codes, which are vulnerable to crashes and malicious attacks. It is crucial to detect and identify bugs

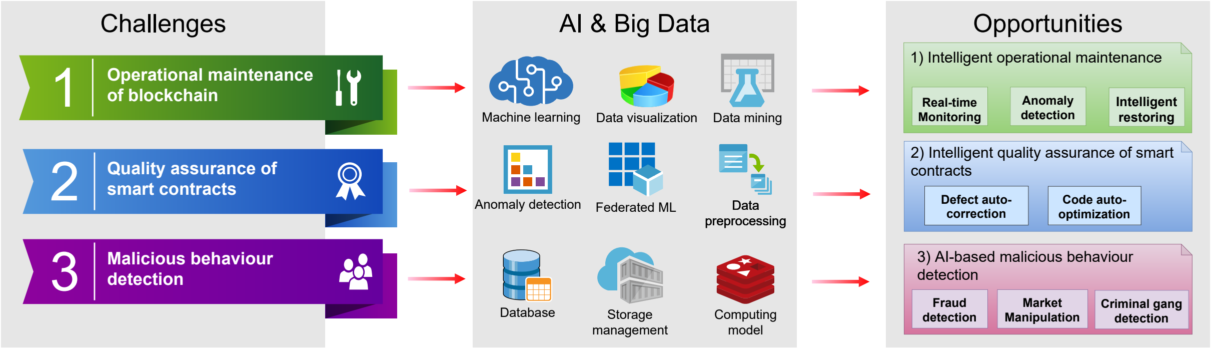


Fig. 2. Opportunities brought by artificial intelligence

in smart contracts so as to achieve the ultimate goal of intelligent quality assurance of smart contracts. The work of [9] presents a symbolic-execution platform namely Oyente to detect and recognize potential bugs. Meanwhile, smart contracts are essentially program codes that are sensitive to execution cost (e.g., the execution of smart contracts is charged by gas in Ethereum). Thus, it is a necessity to identify the gas-costly patterns and correct these vulnerable smart contracts. In [3], Chen et al. developed a tool namely GASPER to identify and locate seven gas-costly patterns via analyzing bytecodes of Ethereum smart contracts. Machine learning methods can be used to detect and recognize vulnerable bugs in smart contracts automatically. Moreover, the growing number of smart contracts also brings the opportunities to automate the composition of multiple contracts. In particular, we can find and identify contracts that fulfill users' requirements and composite them together to realize more comprehensive applications. Different from conventional distributed software systems such as web services [11], smart contracts lack of semantic description and Quality-of-Service (QoS) evaluation metrics. As a result, new AI-based approaches are expected to automate labeling semantics of smart contracts and offer data-driven QoS evaluation of smart contracts.

3) Automated malicious behaviour detection. The decentralized blockchain systems result in the difficulty in auditing malicious behaviours such as money laundering, phishing, gambling and scams that occurred in blockchain platforms. Blockchain systems have generated massive transaction data, which are essentially available to everyone, whereas the historical transaction data are pseudonymous through anonymizing account addresses. The massive blockchain data brings the opportunities in auditing and detecting malicious behaviours. Big data analytics on massive blockchain data can help to identify malicious users, recognize behaviour pattern, analyze market manipulation, detect scams. The work of [12] presents a cross-graph analysis on Ethereum data and identify several major activities occurring on Ethereum blockchain platforms. As in [10], a machine learning based approach was proposed to detect and capture Ponzi schemes that took place in Ethereum. The work of [13] analyzes the leaked transaction history of Mt. Gox Bitcoin exchange and identify a number of market manipulation patterns via singular value decomposition (SVD) method. Moreover, malicious users may exploit multiple anonymous accounts to form a criminal gang to conduct illegal activities on blockchain systems. New machine learning approaches as well as association analysis on multiple accounts are expected to address this issue.

The advances in AI, machine learning and big data analytics bring numerous opportunities to address the aforementioned blockchain challenges. We next present a case study to demonstrate the feasibility of blockchain intelligence.

IV. Case Study

Big data analytics of blockchain data is beneficial to fraud recognition of transactions and vulnerability detection of smart contracts. However, it is also challenging to conduct big data analytics of blockchain data. 1) It is extremely time consuming to download the entire blockchain data due to the bulky blockchain size, e.g., it took more than one week and over 500 GB storage space to fully synchronize (i.e., download) the entire Ethereum at a newly-joined peer. 2) It requires substantial efforts in extracting and processing blockchain data. First, blockchain data is stored at clients in heterogeneous and complex data structures, which cannot be directly analyzed. Meanwhile, the underlying blockchain data is either binary or encrypted. Thus, it is a necessity to extract and process binary and encrypted blockchain data so as to obtain valuable information while this process is non-trivial as conventional data analytic methods may not work for this type of data. 3) There is no general data extract tool for blockchain data. Although several open source tools for blockchain data extraction are available, most of them can only support to extract partial blockchain data (not the entire data).

A. Data Extraction from Ethereum

To address the above challenges, we propose a blockchain data analytics framework namely XBlock-ETH to analyze Etherium data. In particular, we extract raw data consisting of 8,100,000 blocks of Ethereum. Figure 3(a) illustrates the typical Ethereum transaction execution flow from Block N to EVM through blockchain peer. During this procedure, we collect the three types of blockchain raw data: Block, Receipt and Trace. Since the analysis on the raw blockchain data is

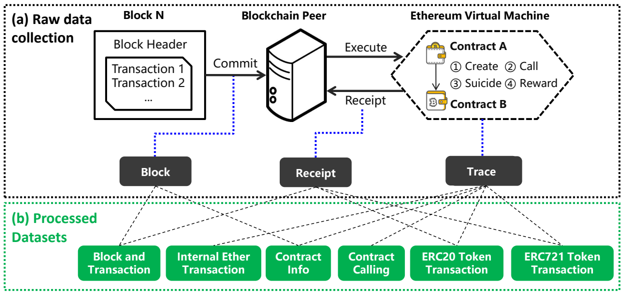
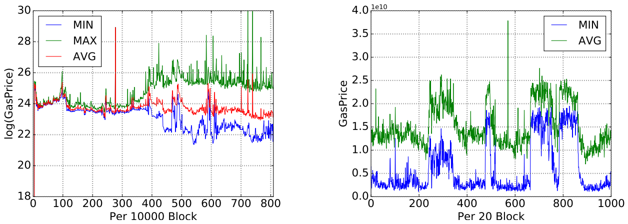


Fig. 3. Data processing during Ethereum transaction flow



(a) Macro view of GasPrice (b) Micro view of GasPrice

Fig. 4. Data visualization of Gas Price of Etherium

difficult, we process and categorize the obtained Etherium Blockchain data into six datasets: *(1) Block and Transaction*, *(2) Internal Ether Transaction*, *(3) Contract Information*, *(4) Contract Calls*, *(5) ERC20 Token Transactions*, *(6) ERC721 Token Transactions* as shown in Figure 3(b). It is non-trivial to process the raw since it requires substantial efforts in extracting useful information from raw data and associating with six datasets.

We then conduct statistic analysis on these refined datasets. In Ethereum, a miner has a higher priority to package the transactions with higher "`gasPrice`" into the block. The visualization of "`gasPrice`" is shown in Figure 4. In a macro view (as shown in Figure 4(a)), the "`gasPrice`" is gradually decreasing with the development of the Ethereum community, except for several peaks caused by extremely frequent transaction when the network is congested. In a micro view (as shown in Figure 4(b)), we extract the time from 8,000,000 to 8,020,000 blocks and find that such fluctuations of "`gasPrice`" can be observed by the tidal law. This observation implies that the fluctuations of "`gasPrice`" can potentially be predicted.

B. Detection Ponzi Schemes from Ethereum

Blockchain can also be exploited to conduct illegal activities such as scams. For example, in [10], we propose a method to detect Ponzi scams in Etherium blockchain through extracting and analyzing key characteristics of user accounts and operation codes in Etherium contracts. Figure 5 makes a comparison between a normal contract and a Ponzi scheme contract in terms of Ether Flow Graph, where the horizontal axis denotes the time line, the vertical axis represents the number of participants in a particular contract, red circles and green circles denote the investment transactions and payment transactions, respectively. In addition, the size of circle also represents the amount of transactions, i.e., the larger circle means the larger amount of transactions.

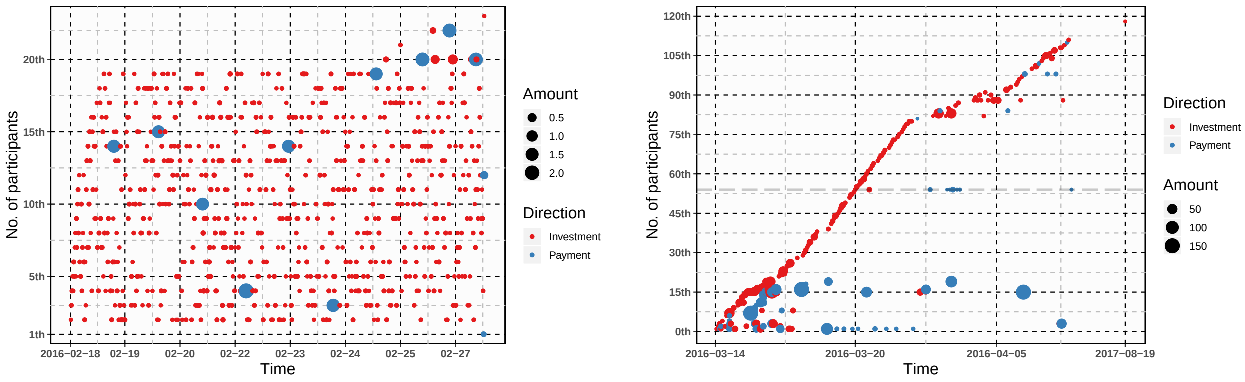
Figure 5(a) is a normal lottery contract while Figure 5(b) shows a typical Ponzi scam (namely Rubixi). We can observe several significant differences between Figure 5(a) and Figure 5(b): 1) there are more participants in Figure 5(b) than Figure 5(a); 2) there are more payment transactions in Figure 5(b) than Figure 5(a) which exhibits more randomness in the number of transactions. After extracting the key features and applying other data mining and machine learning methods, we can successfully classify Ponzi scams from other normal activities.

V. Conclusion and Future Directions

This article first reviews the blockchain technologies and analyze the challenges in blockchain systems. We then introduce artificial intelligence (AI) as well as opportunities brought by AI to blockchain systems. We name such integration of blockchain and AI as *blockchain intelligence*. We mainly discuss that AI bring benefits to blockchain in aspects of *intelligent operational maintenance of blockchain*, *intelligent quality assurance of smart contracts* and *automated malicious behaviour detection*. In addition, we also give a case study to further demonstrate great potentials of blockchain intelligence.

We believe that the integration of AI with blockchain technology will further drive the benignant development of blockchain systems. We outline the future directions in blockchain intelligence as follows.

- *Real-time and automated operational maintenance of blockchain.* Instead of performance monitoring and fault detection, the future intelligent operational blockchain systems are expected to conduct real-time monitoring on multiple performance metrics of blockchain systems and achieve the automated restoring from crash spots.
- *Collective intelligence bestowing smart contracts.* The current QoS assurance approaches of smart contracts are mainly based on the analysis on contract patterns (i.e., detecting and classifying vulnerable contracts) at a sole peer in the blockchain. In contrast to a single intelligent agent, *collective intelligence* can motivate all participants to engage in analyzing and reasoning, thereby contributing their collective knowledge to make better decision in a global context. In the future, collective intelligence is expected to integrate with decentralized blockchain system so as to offer a trustworthy and intelligent provision of blockchain services.
- *Integrating multiple machine learning approaches to monitor and supervise blockchain data.* Decentralization of blockchain brings the difficulty in monitoring and supervising transactions in blockchain platforms, consequently resulting in a number of illegal behaviours. Meanwhile, both heterogeneity and pseudonymity of blockchain data make this situation even worse. In the future, multiple machine learning approaches should be



(a) LooneyLottery contract (*i.e.*, normal contract)

(b) Rubixi contract (*i.e.*, Ponzi scam contract)

Fig. 5. Ether Flow Graph of two smart contracts [10].

integrated together to extract key features from different types of blockchain data. In addition, a dynamic graph (or network) of transactions is also expected to identify the association between different accounts so as to recognize the malicious behaviours.

References

[1] X. Li, P. Jiang, T. Chen, X. Luo, Q. Wen, "A survey on the security of blockchain systems", *Future Generation Computer Systems*, 2017, https://doi.org/10.1016/j.future.2017.08.020.

[2] N. Szabo, "The idea of smart contracts", *Nick Szabo's Papers and Concise Tutorials*. http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart_contracts_2.html

[3] T. Chen, X. Li, X. Luo, X. Zhang, "Under-optimized smart contracts devour your money", *Proceedings of 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 2017, pp. 442–446.

[4] J. Bonneau, J. Clark, S. Goldfeder, "On bitcoin as a public randomness source", *IACR Cryptology ePrint Archive*, 1015, 2015

[5] The DAO, The Hack, The Soft Fork and The Hard Fork (2017). https://www.cryptocompare.com/coins/guides/the-dao-the-hack-the-soft-fork-and-the-hard-fork/

[6] P. Zheng, Z. Zheng, X. Luo, X. Chen, X. Liu, "A detailed and real-time performance monitoring framework for blockchain systems," *2018 IEEE/ACM 40th International Conference on Software Engineering: Software Engineering in Practice Track*, pp. 134–143, 2018. (conference proceedings)

[7] H.-N. Dai et al., "Big data analytics for large-scale wireless networks: Challenges and opportunities,", *ACM Computing Surveys*, vol. 52, no. 5, 2019

[8] T. T. A. Dinh, R. Liu, M. Zhang, G. Chen, B. C. Ooi and J. Wang, "Untangling Blockchain: A Data Processing View of Blockchain Systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 7, pp. 1366-1385, 1 July 2018

[9] L. Luu, D.-H. Chu, H. Olickel, P. Saxena, and A. Hobor, "Making Smart Contracts Smarter,"*Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (CCS '16), 2016

[10] W. Chen, Z. Zheng, J. Cui, E. Ngai, P. Zheng, and Y. Zhou, "Detecting Ponzi Schemes on Ethereum: Towards Healthier Blockchain Technology," *Proceedings of the 2018 World Wide Web Conference* (WWW '18), 1409-1418.

[11] P. Rodriguez-Mier, C. Pedrinaci, M. Lama and M. Mucientes, "An Integrated Semantic Web Service Discovery and Composition Framework," *IEEE Transactions on Services Computing*, vol. 9, no. 4, pp. 537-550, 1 July-Aug. 2016.

[12] T. Chen et al., "Understanding Ethereum via Graph Analysis,"*IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2018, pp. 1484-1492.

[13] W. Chen et al., "Market Manipulation of Bitcoin: Evidence from Mining the Mt. Gox Transaction Network",*IEEE Conference on Computer Communications*, 2019, pp. 964-972.