Research on the Construction of Social Governance Intelligent System in the Era of Big Data

YuChun Li[1,2]

[1,2]e-mail: 523967155@qq.com

[1]Department of Public Administration, Dongguan City College, Dongguan, China

[2]Department of Public Policy and Management, Shih hsin University, Taipei, China

Abstract. With the advent of the era of intelligence and data, the research on public governance has undergone profound changes in theory and practice. As a factor of production, data plays an increasingly key role in social governance. How to realize secure query, collaborative management and intelligent analysis of multi-party data is the key to improve the effect of social governance. Based on secure multi-party computing, blockchain technology and precise intelligence theory, a social governance intelligent system based on big data is proposed. The proposed system can support various applications of social governance and provide decision support for the improvement of social governance level in the new era.

Keywords-Big data, Social governance, Governance model

1 INTRODUCTION

With the rapid development of information technology, various government functional departments have accumulated large-scale data. "Technology support" is an important part of the social governance system, and the new generation of information technologies such as big data, blockchain, cloud computing, and artificial intelligence will provide key scientific and technological support for social governance. At present, in my country's social governance, the government is the leading force in social governance, but it is not the only participant. Enterprises and institutions, social organizations, and urban and rural community residents' organizations have all become important in social governance. These diverse participants together constitute the governance unit in social governance. However, different governance units often store multiple types of massive data, and the management of data rights is also complicated, which is not conducive to efficient data management in social governance, intelligent analysis brings challenges.

2 SOCIAL GOVERNANCE INTELLIGENT FRAMEWORK

The distributed social governance model system includes a total of 3 layers, which are the security computing layer, the blockchain layer, and the social governance layer from the bottom to the top. First, in the secure computing layer, for multi-type massive data, the system performs distributed multi-node storage, and realizes multi-party secure data query through

ICIDC 2022, June 17-19, Qingdao, People's Republic of China

DOI 10.4108/eai.17-6-2022.2322622

secure multi-party computing technology; on this basis, at the blockchain layer, through the shared access control of the blockchain, the one-stop management of data access rights is realized, and the intelligent algorithm of the social governance layer is supported; finally, in the social governance layer, the intelligent decision-making of social governance is realized by combining the construction theory of complex network model and the precise intelligence theory. The specific functions of each layer are as follows.

2.1 Secure Computing Layer

The security computing layer of the system is mainly responsible for designing a unified security basic operator on the basis of distributed storage data, and implementing an efficient multi-party query interface, providing data support for the confirmation of rights of the blockchain layer and the intelligent decision-making of the social governance layer.

2.1.1 Basic Operators for Secure Computing

How to complete the data query in the data distributed storage of different governance units under the restriction of privacy protection is the primary issue of social governance. Secure multi-party computing is a cryptographic technology that can ensure the security of the multi-party data computing process.

•Secure summation. The goal of secure summation is to sum the data of multiple parties, and at the same time ensure that during the calculation process, the data of each party will not be leaked to any other party. This system realizes this operation through secret sharing[1]: first , each party randomly generates a polynomial of degree $n-1$, and sets the locally stored data as the constant term of the polynomial; then, each party substitutes n specific values into the polynomial to calculate the result; finally, in the polynomial result of each party, the summation result can be calculated by the lagrangian interpolation formula on the basis . Ensuring data security, the original data of all parties does not leave the local in this process.

•Security comparison.The goal of security comparison is to sum the data of multiple parties, and output the relationship between the summation result and the given value, and at the same time ensure that the data of each party will not be leaked to any other party during the calculation process. Assuming the data stored in multiple parties is a_i, and the given comparison value is b. First, we divide the comparison value b into n equal parts. From the definition of c_i, we can find that the comparison result can be obtained only by judging whether$\sum_i c_i$ is a positive number. In order to ensure the security of the calculation process, each party generates a random positive number d_i , and calculates$(\sum_i c_i)(\sum_i d_i)$ through the secret sharing protocol, and the positive or negative of $\sum_i c_i$ can be judged.

2.1.2 Multi-party query interface

On the basis of secure computing operators, in order to provide data interfaces for data analysis in social governance, the system implements a series of basic data query operations.

• Multi-party range query. The input of the multi-party range query is the query range, and the output is the data of each party within the query range. In order to ensure that the private data of each party is not leaked during the range query process, the query mechanism is completed by the security set summation operator: first, for a given query range, each party performs a range query locally to obtain its own query result; then, the security set union operator is used to calculate the union of the query result set of multiple parties.

• Multi-party neighbor query. The input of the multi-party neighbor query is the query data and a positive integer k, and the output is the k data that are closest to the query data in the data of each party. Two basic operators for safe set merging are completed: first, according to the given query data, a random radius is generated to obtain the initial query range; then, based on the initial query range, a multi-party range query is performed to obtain the query result set. Compare the size of the query result set with the size relationship of k through the safe comparison operator :if the number of data in the query result set is greater than k, the query range is reduced to half and the range query is performed again; if the number of data in the query result set is less than k, the query range is expanded to 2 times, and performed the range query again; otherwise, the number of data in the query result set is equal to k, and the query result set is directly output . Through the above iterative algorithm, the multi-party nearest neighbor query can be completed more efficiently.

2.2 Blockchain layers

The blockchain layer of the system is mainly responsible for establishing a blockchain system on the query interface provided by the secure multi-party computing technology to achieve right confirmation control.

2.2.1 Shared Access Control Mechanism

The shared access control mechanism is mainly divided into two modules: access control policy formulation and access control automatic authorization.

• Access control policy formulation. This paper proposes feature-based access control . The policy generation algorithm generates access control policies by unifying the relevant attributes in the social governance system, and designs efficient algorithms to quickly search for access control policies. When a governance unit U (such as a neighborhood committee) initiates a request to access a certain data D_i, it can be summarized into four-dimensional attributes A(U)=<$attr_i$,$attr_d$,$attr_a$,$attr_e$>, which respectively represent the originator attribute, access data attribute (data unit, data description,etc),operation attribute and environmental attribute of this visit. The data host judges access control through the attribute information A(U) that initiates the access.

• Automatic authorization of access control. The smart contract that realizes automatic execution should take the initiation of the access request as the triggering condition. Two steps are required in the execution of the contract: one is to find the corresponding access control policy set according to the data attribute A (U) of the access request; the second is to calculate the intersection of attribute requirements in the policy and access request attributes to determine the authorization result. Based on the attributes and policies in the blockchain and smart contracts, the trusted access control of the system can be completed, and the right confirmation control can be achieved[2].

2.2.2 Efficiency optimization of blockchain system

This paper proposes an optimization scheme for system efficiency:

- Node partitioning.The blockchain nodes are divided into two categories: full nodes and light nodes. The full nodes are responsible for collecting and storing the calculation result data in the entire blockchain system, and generating data summaries through the authentication data structure; in social governance, a dedicated social governance data node can be established as a full node. The light node is a distributed governance unit in the social governance system, which only stores the block header of the latest block data of the blockchain, and the block header contains the authentication data of the full node. The digest value is generated by the structure. Authentication data structure is a common technology in outsourcing database[1], which can realize the user's verification of data query results.

- Index structure design. This paper builds an efficient query index for hierarchical storage: first, it builds a hierarchical index according to the matching relationship between global schema and local schema attributes; then, the associated Merkle values are generated for them respectively to realize the credible verification of the index . Among them, the global schema also includes the Merkle tree root, and the Merkle value is determined by the corresponding localv. The Merkle value of the mode is obtained by cryptographic hash functions such as SHA-2; the local mode stores the intermediate node of the Merkle tree, and the Merkle value is obtained by the cryptographic hash function of the Merkle values of all its local mode attributes; the Merkle values for local schema properties are generated by the local schema owner. The index owner can check whether the index has been tampered with by comparing the Merkle value of the child node with the Merkle value of the parent node in the index after passing through the hash function, so as to ensure the credibility of the index.

2.3 Social governance layer

On the basis of the shared access control provided by the blockchain layer,the social governance layer of the system is mainly responsible for realizing intelligent decision-making of social governance through precise intelligent technology. In social governance, the important issue of intelligent decision-making is the complex behavior of social system participants modeling and prediction.

2.3.1 Building complex network models

The contemporary social system can be regarded as a complex information system formed by independent individuals and the connections between individuals through contacts, social connections, and the internet. This actual complex information system usually has the characteristics of huge scale and dynamic laws. Information diffusion is a typical typical case in social governance. The characteristics of randomness are developed, and a dynamic mathematical model is established. Information dissemination is an important dynamic process in complex systems, and it is also a key component of a distributed social collaborative governance system.

Information dissemination is an important dynamic process in complex systems and a key component of a distributed social collaborative governance system. This information dissemination includes top-down dissemination and self-organizing dissemination.The complex system of Behavior characteristics are essentially to study the random nonlinear relationship between system elements, establish a dynamic mathematical model, and reveal the global evolution law of the system, so as to achieve the purpose of efficient, distributed and cooperative control of dynamic systems. For the dynamic process of complex systems, generally can be abstracted into a group consisting of several disjoint groups representing individual state sets. Due to the continuous random changes in the states of individual members, the number of individuals in the group can be characterized by the following principles:

$$\partial_\tau X^{[m]} = \sum {v_{hg}}^{m} \ a_{h,g} N^{-1} X^{[h]} X^{[g]} + \sum {v_h}^{m} a_h X^{[h]} . \qquad (1)$$

where, $X^{[m]}$ is the number of individuals in state m; $a_{h,g}$ and a_h are the conversion rates between dynamic process groups; ${v_{hg}}^{m}$ and ${v_h}^{m}$ take the value of 1, 0 or 1, which is used to represent the number of changes in group m individuals generated according to the coupling interaction between groups; N is the total number of individuals in the system.(1) This type of system includes first-order linearity between groups relationship and second-order nonlinear relationship;the difficulty lies in how to establish the analytical coupling relationship expression and how to fit the parameters in the dynamic equation. To determine the coupling relationship between groups, the mathematical method of the dynamic system can be used to predict the complex. The periodic orbit and steady state of the system can be obtained from the initial state of the system, and the dynamic statistical law of the group can be obtained. After the system parameter set is determined, the predictability and chaotic effect of the system can be further analyzed, and the precise inherent law of system evolution can be obtained. Governance provides a theoretical model.

2.3.2 Precise and intelligent technology application

The following takes the problem of finding super-spreaders in epidemic prevention and control as an example to introduce the application of precision intelligence in social governance. In the spread model of complex networks, the SIR spread process on heterogeneous networks is a classic model. In the propagation model, individuals in the network are divided into three categories, namely susceptible, infected and recovered. In each unit of time, the infected has a certain probability to infect a neighbor node. In the information dissemination model, the individuals in the network can be defined as corresponding groups, that is, the information unknown, the information disseminator and the information averse. In each unit of time , the communicator has a certain probability to spread the information to an unknown of a neighbor node, and make it an information communicator. And the communicator has a probability to become an information averse and stop spreading information. On this basis, we consider a typical multi-dimensional independent communicator model.The dynamic model of this kind of problem can be expressed as

$$\frac{\mathrm{d} i_k(t)}{\mathrm{d} t} = -\mu i_k(t) + \beta k s_k(t) \theta_k(t) + \sigma(t) f(i_k(t), s_k(t)), \qquad (2)$$

where i_k (t) is the ratio of infected persons with degree k to the total population; s_k (t) is the ratio of susceptible persons; The ratio of nodes with k to the total nodes. This takes into account that due to the uneven distribution of node degrees in the network.(2)The probability of reaching a specified node from a random node along an edge in the network is different, proportional to the degree of the node.

In real social governance, the topology structure of the network presents a dynamic evolution law. We need to fully consider the dynamic characteristics of the structure, and modify and improve the traditional model to achieve the goal of intelligence, accurate prediction and social governance.

For example, we consider the time-varying generator function of a complex network:

$$G(x,t) = \sum_{k=0} p_k(t)x^k \qquad (3)$$

Among them, $p_k(t)$ is the ratio of nodes with time degree k at t to the total nodes.(3)

For example, dynamic home isolation during a disease pandemic, reduces the average node degree and the degree of core nodes in a complex system can effectively ,and controls the speed and scope of disease transmission. The above expression paradigm can describe this kind of Dynamic network degree distribution, which in turn improves the accuracy of system control and prediction.

On the basis of accurately constructing the topology structure of complex information system, and finding the central node of complex network, that is, super-spreader, is a key method to control the dynamic behavior of system members according to the network topology. Among them, the traditional definition of the centrality of network nodes is as follows: there are also newer methods such as PageRank, Collective Influence (CI), High Degree Adaptive (HDA). PageRank can give the value of the centrality of each node; although the CI and HDA methods cannot give the value of the centrality of the node, they can give a ranking reflecting the centrality of the node. Among them, the HDA method is a class based on the degree distribution The method of effectively finding super-spreaders is more concise and efficient than CI. The pseudo-code of the HDA method is shown in Algorithm 1.

Algorithm 1. HDA method.

Input: G: directed graph.

Output: L: ordered list of nodes.

```
number of nodes in N   G
for $1 \le i \le N$ do
    $v_{\max} \leftarrow \varnothing$
    for $v'$ in $G$ do
        if $v'.degree > v_{\max}.degree$ then
            $v_{\max} \leftarrow v'$
Remove the node $v_{\max}$ and all its connected edges, and update the graph G
The centrality rank of node v is set to i
Add nodes to the ordered list L in order of centrality ranking
return L
```

Different from the degree centrality, the HDA method ensures that every node found is the node with the largest degree in the subgraph after excluding the node with higher centrality ranking, which makes this method absolutely superior to the degree center in finding super-spreaders. In the actual complex network, the HDA method is also close to the results of the existing optimal methods.

3 SYSTEM EXPERIMENTAL VERIFICATION

Experiments are carried out to verify the basic operation of secure multi-party computing and the effect of precise intelligent governance. The secure computing layer of the system can realize efficient data query, such as finding close contacts in a specific area and a specific time range, helping people in epidemic prevention and control troubleshooting. First, the range query and nearest neighbor query are taken as examples to verify the query efficiency and communication overhead of the secure computing layer.

3.1 Experimental setup

3.1.1 lab environment

This paper builds a secure multi-party computing scenario through 5 servers .The server operating system is Ubuntu 18.04.5 LTS, the memory is 64 GB, and it is equipped with a 32-core Intel(R) Xeon(R) Gold 5118 2.30 GH CPU. Four servers simulate different levels of multi-party governance units by running multiple processes, and one server acts as the application side of governance to perform query operations by integrating multi-party data. The experiment is carried out in a network environment with a bandwidth of 10 GB/s.

3.1.2 Experimental data

In this experiment, the trajectory data of 15 taxi companies in Guangzhou is used, and the data set contains the trajectory data of 1138,161 taxis for 2 months.

3.1.3 Comparison Algorithms

This system implements multi-party query operations based on the Secure Query Operator in Hu-Fu[3] . The baseline query algorithms compared in the experiments include plaintext query algorithms, the recent secure query system Conclave and SMCQL [4].

■ Plaintext query. The plaintext query does not use any security encryption technology, and directly aggregates the plaintext query results of multiple parties. Although the plaintext query algorithm is difficult to meet the data security requirements, it can be used as the optimal value for the computational efficiency and communication overhead of the query algorithm.

■ Conclave system. Conclave is a secure multi-party computing technology based on secret sharing, which implements basic secure query operations based on Sharemind[5]. This experiment applies it to multi-party range query and multi-party nearest neighbor query and compares them.

■ SMCQL system. SMCQL is a secure computing primitive that can convert SQL query primitives into ObliVM [6]. On this basis, multi-party secure queries are implemented, and then secure aggregation operations are realized . Since ObliVM only supports two-party operations, the SMCQL system Only two-party queries are supported. This experiment applies and compares multi-party range queries and multi-party nearest neighbor queries.

3.1.4 Experimental parameters

In the multi-party range query and the multi-party neighbor query, the default number of participants is 6.The variation parameters are shown in Table 1. In the multi-party range query, the query scope and data size are changed respectively. The bold parameters are the default parameters; in the multi-party neighbour query, the query parameters are changed respectively. Number of neighbors and data size . Since SMCQL only supports two-party experiments, it is compared separately in the scenario, where the governance unit is two-party.

Table 1 Experimental parameters

Action name	Parameter name	Parameter range
Multi-party range query	Query scope	10^{-5}%, 10^{-4}%, 10^{-3}%, 10^{-2}%, 10^{-1}%
	Data scale	10^4,10^5 ,10^6,10^7, 10^8
Multi-party nearest neighbor query	Number of neighbors	4, 8, 16, 32, 64
	Data scale	104,105 ,106,107, 108

3.2 Multi-party range query

3.2.1 Change query scope

First, we analyze the running time of changing the query range to calculate the multi-party range query. Compared with the Conclave system, the running time of the secure query operator implemented by this system is significantly shorter, and the multi-party range query calculation time of this system is about 43% of the Conclave system . The calculation time of the system in multi-party range query is 3.4 times that of plaintext calculation, while the Conclave system is 8.8 times that of plaintext calculation . Secondly, the communication overhead under different query ranges is analyzed . The communication cost of multi-party range query in this system is 4.2 times that of plaintext calculation . And the communication overhead of Conclave can reach 69 times that of plaintext calculation. Therefore, this system has obvious advantages over Conclave in terms of query efficiency and communication overhead .

3.2.2 Change data size

The experimental results under different data scales are similar to the experimental results of different query ranges . The system is superior to the Conclave system in terms of query efficiency and communication overhead. In terms of query efficiency, the running time of this system is 38.8%–40.9% of that of the Conclave system; In terms of communication overhead, the advantages of this system are more obvious, and the communication overhead is only 4.4% of the Conclave system. Compared with the plaintext calculation, the communication overhead increases by an average of 3.3 times, which is within an acceptable range .

3.2.3 Compare with SMCQL

This system is compared with SMCQL when the number of participants is two, and the default parameters are selected for other parameters.As can be seen from Table 2, In the multi-party range query, this system has obvious advantages in running time and communication overhead compared with the SMCQL system. The system-wide query time isSMCQL 31.3% of SMCQL, while the communication overhead is only 63% of SMCQL .

Table 2 SMCQL comparison results

Name	Parameter name	Running time(ms)	Communication overhead(KB)
Plaintext calculation	Range query	25.05	2.88
	Neighbor query	21.82	9.43
SMCQL	Range query	425.07	56.61
	Neighbor query	1604.21	2072
This system	Range query	133.17	35.29
	Neighbor query	26.07	39.41

3.3 Multi-party neighbor query

3.3.1 Change the number of neighbors

Firstly, the security query time of each system under different number of neighbors is analyzed . it can be seen that when the number of neighbors is small, such as when the number of neighbors is 2, the query time of the Conclave system is less than that of this system; however, with the increase of the number of neighbors, the query time of the Conclave system increases rapidly, and the query efficiency of this system has obvious advantages . When the number of query neighbors is 8, the query time of this system is 68.3% of that of Conclave; When the number of query neighbors is 64, the running time of the multi-party neighbor query of the Conclave system is 23.7 times that of this system. In terms of communication overhead, the communication overhead required for the multi-party neighbors query of the Conclave system is huge, up to 55 656 KB, which is beyond the acceptable range; while the communication overhead required by the plaintext calculation and this system is less than 100 KB, which is within the acceptable range.

3.3.2 Change data size

Changing the data size has a limited impact on the system running time and communication overhead. The running time of this system is close to theplaintext calculation, which is only 23.9% of the Conclave system. In terms of communication overhead, the cost of executing multi-party neighbor query in Conclave system is huge, which is 1 600 times and 6 500 times of the computing cost of this system and plaintext respectively . The communication overhead of this system is less than 100 KB, which is acceptable.

3.3.3 Compare with SMCQL

In both scenarios, the system is superior to the SMCQL system in terms of query efficiency and communication overhead. The running time of SMCQL is 61.7 times that of the system, and the communication overhead is 53 times that of the system . The plaintext calculation is in the same order of magnitude, with obvious advantages.

4 CONCLUSIONS

This paper proposes a distributed social intelligent governance system based on big data to deal with the challenges brought by the complex problems of social governance. The system implements secure computing basic operators through secure multi-party computing technology, and provides a multi-party secure query interface for social governance; it builds a decentralized spatial structure based on blockchain technology to provide data platform support for efficient processing of emergencies; it builds non-linear models through precise intelligence, and builds interpretable and adjustable artificial intelligence for system behavior evolution and global dynamic analysis. It provides scientific decision support for nonlinear factors in social governance . The social governance intelligent system proposed in this paper will further support various applications of social governance including public health security, smart transportation, and provide a platform and a platform for the improvement of social governance in the new era.

Acknowledgments. This paper is the phased achievement supported by the young teacher development fund project of Dongguan City college "Research on the scientific model and path of social governance from the perspective of rule of law" (Project No.: 2020qjy002r); The phased achievements is funded by the college students' innovation and entrepreneurship training project "co-construction, co-governance and sharing: Exploration and Path Research of social organizations participating in urban community governance" (Project No.: 202013844056) .

REFERENCES

[1] Beimel A. Secret-sharing schemes(2011): A survey. In: Proc. of the 2011 Int'l Conf. on Coding and Cryptology. Berlin, Heidelberg:Springer, 11−46

[2] Jin Guantao(2017). Reflecting on the artificial intelligence revolution . Social Science Digest, 11,12-26.

[3] Miller A, Hicks M, E(2014). Authenticated data structures, generically. In: Proc. of the Symp. on Principles of Programming Languages. San Diego: ACM, 411−424.

[4] Tong YX, Pan XC, Zeng YX. Hu-Fu(2021): Efficient and secure spatial queries over data federation. 2021. https://github.com/BUAA-BDA/Hu-Fu.

[5] Beimel A. Secret-sharing schemes(2011): A survey. In: Proc. of the 2011 Int'l Conf. on Coding and Cryptology. Berlin, Heidelberg:Springer, 2011. 11−46.

[6] Liu C, Wang XS, Nayak K, Huang Y, Shi E. Oblivm(2019): A programming framework for secure computation. In: Proc. of the 2019 IEEE Symp. on Security and Privacy. IEEE, 2019. 359−376.