OIDPR: Optimized Insulin Dosage based on Privacy-Preserving Reinforcement Learning

1st Zuobin Ying
School of Electrical & Electronic Engineering
Nanyang Technological University
School of Computer Science & Technology
Anhui University
Singapore
yingzb@ahu.edu.cn

2nd Yun Zhang
School of Computer Science & Technology
Anhui University
Hefei, China
1417178632@qq.com

3rd Shuanglong Cao
School of Computer Science & Technology
Anhui University
Hefei, China
610237563@qq.com

4th Shengmin Xu
School of Information Systems
Singapore Management University
Singapore
smxu1989@gmail.com

5th Ximeng Liu
College of Mathematics & Computer Science
Fuzhou University
Fuzhou, China
xmliu@smu.edu.sg

***Abstract*—The precise dosage of insulin plays an important role in the treatment of diabetes. To offer accurate dosage, some AI-based auxiliary dosing systems have been proposed. Unfortunately, these schemes demand real-time health data, which is highly relevant to the health situation of the diabetics. The traditional personalized drug delivery frameworks for accurate dosing of insulin always collect and transmit medical data in plaintext, which may cause the disclosure of user privacy. Therefore, to optimize insulin dosage and protect privacy simultaneously, we propose a framework for an optimized insulin dosage via privacy-preserving reinforcement learning for diabetics (OIDPR). In OIDPR, both the additive secret sharing and edge computing are deployed to complete data encryption and improve efficiency. The user's medical data is divided into secret shares uniformly at random, then compute separately at the edge servers. During the computation task of Q-learning, data is stored in the format of ciphertext and processed using the proposed additive secret sharing protocol. Finally, comprehensive theoretical analyses and experiment results demonstrate the security and efficiency of our framework.**

***Index Terms*—additive secret sharing, privacy-preserving, individualization dosing delivery**

I. Introduction

Diabetes is a chronic disease world widely, the number of diabetics will increase to 693 million by 2045 [1]. To alleviate the worsening form of diabetes, the species of medicines are increasing correspondingly, which always involving different mechanisms of effect and safety. Multitudinous researches are focusing on the best ways to develop new therapies and optimize prescribing by steering patients to the right drug at the correct dose [2]. Insulin requirements are strictly defined, however, in actual medical diagnosis, diabetic drug management involves complex investigation and coordination of care by a myriad of medical specialists. A clearer understanding of these dynamics highlights the significance of accurate dosing in the medical scenario and in healthcare aimed at improving patient safety. The concept of using machine learning to give the most appropriate drug distribution for each patient's condition is proposed.

In recent years, machine learning has promoted the development of intelligent medicine [3]. Data storage faces the risk of information leakage in the cloud server, which is the main stumbling block that hinders the popularization of the individualized dosing system [4]. The leaked data may be used to infer personal living conditions, place of residence, and even identity information, which could be used to re-identify a person. Moreover, an adversary can use this information for commercial or criminal purposes to gain improper benefits. If these drawbacks are not addressed, the medical community is unwilling to adopt machine learning as a service platform, this puts the situation in dilemma. To popularize this new paradigm, the patient's personal health information should be reserved to prevent unauthorized disclosure by the medical service provider.

The existing medical data privacy protection mainly depends on the following technologies. Traditional anonymous technology, such as *k-anonymity* may not be suitable for medical data desensitization. If k users are in the same location or a sensitive area, such as a hospital, their location information may also be leaked. Another method to preserve privacy is homomorphic encryption (HE), which enables the decryption party can only obtain the final result, without obtaining the message of each ciphertext [5]. Whereas, the feature of expensive complexity and intensive memory consumption make it unpractical in real-world applications. Accordingly, an error-free and efficient framework to address the privacy problem of personalized drug dosage needs to be constructed.

To conquer the difficulty of applying privacy-preserving

Annex to ISBN 978-3-903176-28-7

in real-time scenario, we re-construct the $\mathcal{Q}$-Learning by integrating the secret sharing scheme, in which data is stored in the ciphertext and processed using additive secret sharing. Accordingly, we put forward an optimized insulin dosage via privacy-preserving reinforcement learning, namely OIDPR.

II. Related Work

Machine learning has been widely used in medical, industrial, and national defense research. Among them in recent years, medical research is undergoing a transformation from a "one-size-fits-all" strategy to a precision medical method [6]. Since the individual response to treatment varies among patient populations, due to the prolonged nature of the treatment, patient response may change over time, machine learning can make accurate treatment plans for patients at the right time.

In the era of big data, the value of personal data has received more attention [7]. How to resolve the contradiction between the development of data value and personal privacy protection is an urgent issue. Luo *et al.* [8] propose a practical framework called Privacy Protector, and design a distributed database composed of multiple cloud servers in this framework. Yang *et al.* [9] propose to innovatively combine statistical analysis and cryptography to provide multiple examples of the balance between medical data utilization and privacy protection. In order to solve the privacy leakage problem of outsourcing, Liu *et al.* [10] use HE to design a privacy protection RL framework Preyer. Unfortunately, the computing power and storage space required by HE-based methods, including Preyer, is vast, but also achieve unprecedented success in many challenging areas.

III. Approach

This section summarizes the important algorithm $\mathcal{Q}$-Learning in reinforcement learning, which is the cornerstone of the individualized dosing policy in OIDPR.

A. $\mathcal{Q}$-Learning

$\mathcal{Q}$-Learning is an extensive machine learning model that can recommend optimal strategies for individualized drug dosages for patients with diabetes. For a standard $\mathcal{Q}$-Learning model, there are three entities, an agent, a state space set $\mathcal{S}$, an action space set $\mathcal{A}$, maximization total reward. $\mathcal{Q}$-Learning tries to optimize the agent action selection for each state by virtue of the $\mathcal{Q}$-function $\mathcal{Q}(s_i, a_i)$, where $s_i \in \mathcal{S}$ and $a_i \in \mathcal{A}$. The $\mathcal{Q}$-Value update of $\mathcal{Q}$-Learning is as follows

$$\mathcal{Q}^{new}(s_i, a_i) = \mathcal{Q}(s_i, a_i) + \alpha[r_{i+1} + \gamma \underset{x_{i+1}}{argmax}\mathcal{Q}(s_{i+1}, a_{i+1}) - \mathcal{Q}(s_i, a_i)],$$

where α is the learning rate between 0 and 1, r_{i+1} is the reward after performing action a_i at state s_i, γ is the discount factor. Moreover, ϵ-greedy policy is utilized in $\mathcal{Q}$-Learning to select the action for the current state. The selector uses currently available knowledge to compute

$$a_i = \begin{cases} \text{argmax}_{a'}\mathcal{Q}(s_i, a') & 1 - \epsilon_i \\ \text{randomly select from } \mathcal{A} & \epsilon_i \end{cases}$$

where ϵ_i is the probability for exploration at iteration i, the value of which is set to 1 at the beginning and decreases along with training.

IV. Secret $\mathcal{Q}$-Learning Framework

In OIDPR, *HCP* tries to give a precise dose to the diabetics on the edge server through the deployed $\mathcal{Q}$-Learning model. The details of the OIDPR workflow are shown below.

1) Secure $\mathcal{Q}$-Learning Model Initialization: To build OIDPR, *HCP* first defines finite state set $\mathcal{S} = \{s_1, s_2, ..., s_\delta\}$ and action set $\mathcal{A} = \{a_1, a_2, ..., a_\sigma\}$. $\mathcal{S}$ describes the state space of diabetes data attributes. And $\mathcal{A}$ is related to possible actions that *HCP* may operate. Corresponding to the states and actions, a $\mathcal{Q}$-Table $\mathcal{Q} = \{(s_i, a_j, \mathcal{Q}(s_i, a_j))|s_i \in \mathcal{S}, a_j \in \mathcal{A}\}$ that stores the quality of state-action information is built. The elements of $\mathcal{Q}$ are identically initialized with "0" at the beginning. The three sets are then randomly split into $(\mathcal{S}', \mathcal{S}'')$, $(\mathcal{A}', \mathcal{A}'')$, $(\mathcal{Q}', \mathcal{Q}'')$ and send to ES_1 and ES_2, respectively. The other parameters sent along with them are the learning rate (α', α'') and the discount factor (γ', γ''). In the viewpoints of ES_1 and ES_2, the secret shares are just a mass of random values.

2) Train Data Outsourcing: To train OIDPR, *HCP* collects historical state-action data of the whole diabetics as $\mathcal{H} = H_1 \bigcup H_2 \bigcup ... \bigcup H_\alpha$ according to the time sequence. Based on the definition of $\mathcal{S}$ and $\mathcal{A}$, we can build a very long state-action sequence $\mathcal{N}$ about the diabetics with $\mathcal{H}$. Considering training efficiency, $\mathcal{N}$ is then split into smaller batches $\mathcal{N} = \{N_1, N_2, ..., N_\varrho\}$, where $N_i = \{n_{i,1}, n_{i,2}, ..., n_{i,\tau}\}$, $0 < i < \varrho$, τ corresponds to the time sequence and $n_{i,j}$ = $(s_{i,j}, \langle a_{i,j}\rangle, s_{i,j+1}, \langle r_{i,j+1}\rangle)$. $r_{i,j+1}$ is the reward for the operating action $a_{i,j}$ at state $s_{i,j}$. And $\mathcal{N}$ is randomly split into shares $\mathcal{N}'$ and $\mathcal{N}''$ and send to ES_1 and ES_2 for training. In $\mathcal{N}'$ and $\mathcal{N}''$, where $n'_{i,j} = (s'_{i,j}, \langle a'_{i,j}\rangle, s'_{i,j+1}, \langle r'_{i,j+1}\rangle) \in \mathcal{N}'$, $n''_{i,j} = (s''_{i,j}, \langle a''_{i,j}\rangle, s''_{i,j+1}, \langle r''_{i,j+1}\rangle) \in \mathcal{N}''$, and $n_{i,j} = n'_{i,j} + n''_{i,j}$.

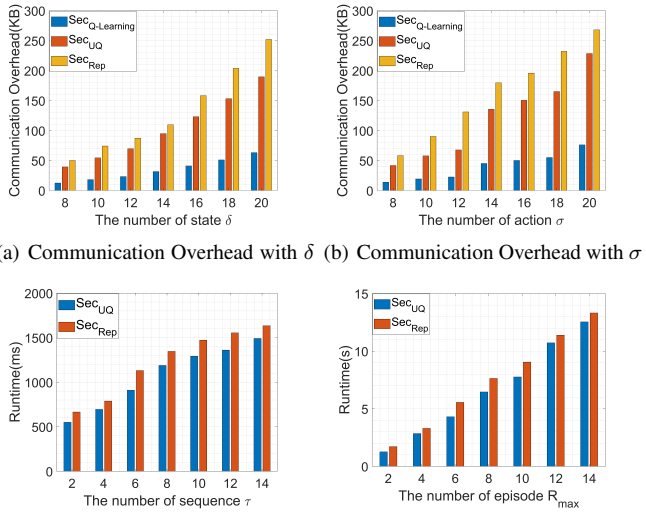
3) Privacy-Preserving Decision Making: To obtain a decision from the trained OIDPR, *UP* splits their current state s_q into uniformly random secret shares (s'_q, s''_q) and sends them to ES_1 and ES_2, respectively. After completing the interactive protocols of OIDPR, ES_1 and ES_2 send back the optimal action decision (a'_q, a''_q), *UP* computes $a_q = a'_q + a''_q$ to recover the plaintext of final output.

V. Performance Evaluation

In this section, comprehensive experiments are operated to prove the efficiency of OIDPR. The experiment data are online available historical data from a diabetes dataset in the UCL machine learning database.

A. Performance Analysis of OIDPR

It can be discovered that four key factors affect the operational efficiency of our protocol, namely, state number δ, action number σ, experience record length τ and number of iteration R_{max}. Therefore, we evaluate the performance changes of the three interaction protocols through four factors. Note that, in the following experiments, the default setting is that the data



(a) Communication Overhead with δ (b) Communication Overhead with σ

(c) Runtime with τ (d) Runtime with R_{max}

Fig. 1. Performance Evaluation of Interactive Protocols

length, ϵ = 0.1, α = 0.1, γ = 0.7, ϱ = 10, τ = 3, and R_{max} = 3, while experimenting the performance change with δ and σ, we set τ = 1 and R_{max} = 1, which correspond to one times $\mathcal{Q}$-Learning computation process cost.

From Fig. 1(a) and Fig. 1(b), it is observed the communication overhead of the interactive protocols Sec_{Uq}, $Sec_{\mathcal{Q}-learning}$, and Sec_{Rep} increase with δ and σ. From Fig. 1(c) to Fig. 1(d), $Sec_{\mathcal{Q}-learning}$ and Sec_{Rep} increase the running time with the augment of τ and R_{max}. Along with the increment of states number, OIDPR needs to call Sec_{Com} and Sec_{Mul} multiple times to locate the target state in the sub-protocols. Therefore, as shown in Fig. 1(a), the interaction messages and communication overhead of the three upper-layer interaction protocols increase correspondingly. As can be seen in Fig. 1(b), the communication overhead also be increased at a similar rate. This is because, according to the further experimental results on the performance of the sub-protocols, it has basically the same effect on the efficiency of the sub-protocols. As can be seen from Fig. 1(c) to Fig. 1(d), the experience memory pool parameters will also increase the calculation and communication costs of OIDPR. Nevertheless, the increase is caused by the increment of invocation times for the basic protocols.

B. Effectiveness Analysis of OIDPR

We compare the efficiency of OIDPR with the homomorphic encryption (HE) based method in [10] in Table II. Here, the setting of key parameters is δ = 10, σ = 10, ϱ= 10, τ = 10, R_{max} = 2. The most important reason for this phenomenon is that, for OIDPR, data encryption and decryption only need to generate a few uniform random values or perform simple addition. However, for HE, a large number of time-consuming exponential operations or other mathematical operations are required.

TABLE I
PROTOCOL RUNTIME AND COMMUNICATION OVERHEAD COMPARISON

	Runtime(s)		**Communication Overhead(KB)**	
	Our Scheme	***[10]***	***Our Scheme***	***[10]***
Sec_{Act}	0.103	31.39	119.1	264.7
Sec_{Ele}	0.224	81.2	238.2	264.7
Sec_{Max}	0.025	0.17	0.9	1.75
Sec_{Gry}	0.140	32.7	123.3	280.4
Sec_{Uq}	0.543	184.8	810.3	1545
$Sec_{\mathcal{Q}-learning}$	1.732	714.1	2231.7	6180
Sec_{Rep}	2.045	1011	3582.3	8734

VI. CONCLUSION

In this paper, we propose a lightweight $\mathcal{Q}$-Learning-based additive secret sharing protocol that can be used in the privacy protection system of personal data of diabetic patients, named OIDPR. This system uses edge servers to reduce model updates and drug dose detection operation completion times. The proposed additive secret sharing makes data encryption and decryption only need additive operations. It reduces the demand for computing power and guarantees efficiency and privacy protection in terms of practicality.

REFERENCES

[1] N. Cho, J. Shaw, S. Karuranga, Y. Huang, J. da Rocha Fernandes, A. Ohlrogge, and B. Malanda, Idf diabetes atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045, Diabetes research and clinical practice, vol. 138, pp. 271C281, 2018.

[2] L. Meneghini, A. Doshi, D. Gouet, T. Vilsboll, K. Begtrup, P.Orsy, M. F. Ranthe, and I. Lingvay, "Insulin degludec/liraglutide (iDeglira) maintains glycaemic control and improves clinical outcomes, regardless of pre-trial insulin dose, in people with type 2 diabetes that is uncontrolled on basal insulin," Diabetic Medicine, 2019.

[3] S. Saria, "Individualized sepsis treatment using reinforcement learning," Nature medicine, vol. 24, no. 11, p. 1641, 2018.

[4] B. Balle, A. Gascon, O. Ohrimenko, M. Raykova, P. Schoppmann, and C. Troncoso, "Ppml'19: Privacy preserving machine learning," in Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, pp. 2717-2718, 2019.

[5] O. Kocabas and T. Soyata, "Towards privacy-preserving medical cloud computing using homomorphic encryption," in Virtual and Mobile Healthcare: Breakthroughs in Research and Practice, pp. 93-125, IGI Global, 2020.

[6] R. Iniesta, D. Stahl, and P. McGuffin, "Machine learning, statistical learning and the future of biological research in psychiatry," Psychological medicine, vol. 46, no. 12, pp. 2455-2465, 2016.

[7] Y. Yang, X. Zheng, W. Guo, X. Liu, and V. Chang, "Privacy-preserving smart iot-based healthcare big data storage and self-adaptive access control system," Information Sciences, vol. 479, pp. 567-592, 2019

[8] E. Luo, M. Z. A. Bhuiyan, G. Wang, M. A. Rahman, J. Wu, and M. Atiquzzaman, "Privacyprotector: Privacy-protected patient data collection in iot-based healthcare systems," IEEE Communications Magazine, vol. 56, no. 2, pp. 163-168, 2018.

[9] J.-J. Yang, J.-Q. Li, and Y. Niu, "A hybrid solution for privacy preserving medical data sharing in the cloud environment," Future Generation Computer Systems, vol. 43, pp. 74-86, 2015.

[10] X. Liu, R. Deng, K.-K. R. Choo, and Y. Yang, "Privacy-preserving reinforcement learning design for patient-centric dynamic treatment regimes," IEEE Transactions on Emerging Topics in Computing, 2019.