Efficient and Privacy-Preserving Decision Tree Classification for Health Monitoring Systems

ABSTRACT

Due to the increasing health care costs and the advance of wireless technology, health monitoring systems have been widely adopted recently. In health monitoring systems, a hospital outsources a clinical decision model to a cloud service provider, which receives biomedical data from remote clients and produces clinical decisions based on the outsourced model. Due to critical privacy concerns, both the clinical decision model and biomedical data should be protected. In this paper, we propose an efficient and privacy-preserving decision tree classification scheme (PPDT) for health monitoring systems. Specifically, we first transform a decision tree classifier (i.e. the clinical decision model) to boolean vectors. Then, we leverage symmetric key encryption to encrypt the boolean vectors as encrypted indexes. The privacy-preserving decision tree classification is achieved by searching the encrypted indexes with encrypted tokens. We formulate a leakage function, and provide security definition and simulation-based proof for PPDT. The performance analyses demonstrate that PPDT is very efficient in terms of computation, communication, and storage. Experimental evaluations show that PPDT only requires microsecond-level execution time, kilobyte level communication costs, and kilobyte-level storage costs on the test dataset.

**EXISTING SYSTEM**

HE-based schemes. Most of the HE-based schemes consider a hospital-client setting for health monitoring systems [10]– [13]. In this setting, the hospital’s clinical decision tree model is required to be protected from the client, while the client’s biomedical features and clinical predictions are required tobe protected from the hospital. HE-based schemes protect the privacy of clinical model and biomedical data by utilizing FHE [10], [11] or AHE [12], [13]. Although HE-based schemes enable privacy-preserving decision tree classification for health monitoring systems, these schemes may face high computational costs due to high-complexity homomorphic operations [10]–[13].

To achieve efficient health monitoring services, it is desirable to design a scheme with low computation costs. Namely, achieving sub-linear computational complexity and avoiding cryptographic tools with expensive computational

costs.

MPC-based schemes. Considering a system model with multiple non-colluding hospitals or multiple non-colluding clients, most of the MPC-based schemes evaluate the prediction interactively and collaboratively. MPC-based schemes protect the privacy of both clinical decision model and biomedical data by utilizing tailored GC [15], SS [16], [19], or hybrid cryptographic tools that combine OT, SS, GC, and

AHE [16]–[18]. Although MPC-based schemes avoid using high-complexity encryption, these schemes are designed based on a multiple party non-collusion security assumption and incur prohibitive communication costs due to collaboratively evaluation. To achieve real-time health monitoring services, it is important to design a novel scheme with low communication costs, i.e., avoiding MPC techniques.

SSE-based schemes. To improve the computational and communication efficiency for health monitoring systems, Liang et al. consider a system model that involves a hospital, a client, and a cloud service provider. In this setting, the hospital outsources the clinical decision model to a cloud service provider, and thus the clinical decision model and biomedical data should be protected against the cloud service provider [20]. Liang et al. extract decision rules from decision tree classifiers, and develop an SSE-based scheme (SDTC) with O(1) computational complexity and 1 round communication interaction [20]. Yet, since the size of indexes in SDTC are exponential to the size of the decision tree classifier, SDTC suffers from prohibitive storage overheads. Furthermore, the large size of indexes increase both the computational cost and the communication cost of SDTC. To achieve practical health monitoring services, it is important to design a scheme with low storage costs. Namely, the storage cost should not be exponential to the size of the input domain.

Disadvantages

* To protect the confidentiality of both clinical decision models and biomedical data, several privacy-preserving decision tree classification schemes.
* MPC-based schemes enable multiple parties jointly and privately classify data according to decision trees, but they may lead to expensive communication costs.

**PROPOSED SYSTEM**

\_ We propose an efficient and privacy-preserving decision tree classification scheme (PPDT) for health monitoring systems. First, we transform decision tree classifiers to boolean vectors, which are indexes that enable O(1) computational complexity for decision tree classification.

With such boolean vectors, PPDT significantly improves computation, communication, and storage efficiency simultaneously. By utilizing symmetric key encryption, pseudo-random functions, and pseudo-random permutations to protect the confidentiality of clinical decision models and biomedical data, PPDT significantly reduces the computational costs due to the adoption of lowcomplexity

cryptographic primitives.

\_ We formulate a security definition and give a simulationbased security proof for PPDT. First, we identify a leakage function L, which includes the size pattern,

search pattern, and access pattern of PPDT. Then, we formulate the L-security definition, which is defined based on the leakage function L. Finally, we provide

a simulation-based security proof to demonstrate that PPDT captures the L-security definition. Namely, both the clinical decision model and biomedical data are well protected.

\_ We conduct performance analyses and evaluations for PPDT. We analyze the computational costs and index sizes of PPDT and the scheme in [20] (SDTC). Despite both PPDT and SDTC are with O(1) computational complexity, the comparison results show that PPDT requires lower computational costs and smaller index sizes than SDTC. The experimental evaluations in Breast-Cancer-Wisconsin dataset also illustrate the performance advantages of PPDT. The performance evaluations demonstrate that: (1) the computational complexity of PPDT is O(1), (2) PPDT only requires micro seconds level execution time, kilobyte-level communication costs, and kilobyte-level storage costs for achieving privacy privacy preserving decision tree classification, and (3) The performance (including computation, communication, and storage efficiency) of PPDT is orders of magnitudes boosted than SDTC.

**Advantages**

1) Data confidentiality. Since biomedical features and clinical decisions are sensitive data for C, the confidentiality of biomedical data should keep secret against CSP.

2) Model confidentiality. Due to intellectual property protection issues, the clinical decision model is valuable knowledge assets for H. Thus, the confidentiality of clinical decision model should be protected against CSP.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows 7 Ultimate.
* **Coding Language :** Python.
* **Front-End :** Python.
* **Back-End :** Django-ORM
* **Designing :** Html, css, javascript.
* **Data Base :** MySQL (WAMP Server).