Chapter

Patient Bayesian Inference: Cloud-Based Healthcare Data Analysis Using Constraint-Based Adaptive Boost Algorithm

Shahid Naseem

Abstract

Cloud-based healthcare data are a form of distributed data over the internet. The internet has become the most vulnerable part of critical healthcare infrastructures. Healthcare data are considered to be sensitive information, which can reveal a lot about a patient. For healthcare data, apart from confidentiality, privacy and protection of data are very sensitive issues. Proactive measures such as early warning are required to reduce the risk of patient's data violation. This chapter investigates the ability of Patient Bayesian Inference (PBI) for network scenario analysis with violation of patient data to produce early warning. The Bayesian inference allows modeling the uncertainties that come with the problem of dealing with missing data, allows integrating data from remote nodes, and explicitly indicates dependence and independence. The use of constraint-based adaptive boost algorithm can demonstrate the patient's Bayesian inference performance in the real-world datasets from healthcare data.

Keywords: Bayesian inference, healthcare, constraint-based learning, explicitly, adaptive

1. Introduction

Healthcare data have always been considered to be sensitive information, which can reveal a lot about a patient. This is why medical confidentiality prohibits a medical professional to disclose information about a patient's case. If a physician does not have accurate information on a patient's health, it may lead to an inaccurate diagnosis and improper treatment. Data concerning health mean personal data related to the physical or mental health of patients, including the provision of healthcare, which are real information about patient's health. Sensitive data concerning health require additional protection as it can go to the core of a human being. Healthcare data come within a person's most intimate sphere. Unauthorized disclosure may lead to various forms of discrimination and violation of fundamental rights. The risk of data processing generally does not depend on the contents of the data but on the context in which they are used [1].

The processing of healthcare data is likely to lead violation of individual rights and interests. Patients' data, which are, by their nature, particularly sensitive in

relation to fundamental rights and freedoms. Data processing could create significant risk to the patient's rights and freedoms. In principle, processing of sensitive data is prohibited, unless a suitable safeguard method is used to protect the data [2]. Derogating from the prohibition to process special categories of a patient data including health data is allowed with the following cases [3]:

- Explicit consent is given by the data subject.
- Processing is necessary to protect the vital interest of a patient if this patient is physically or legally incapable to give consent, for example, in emergency situations or with minors.
- Processing is necessary in order to provide healthcare if the data are processed by or under the responsibility of a professional subject to the obligation of professional secrecy.

2. Risks in cloud-based healthcare data

Cloud computing has many risks related to data confidentiality and data security. The data stored in the cloud are highly confidential, such as patient records. Most of time, data being stored or processed in cloud are in large numbers, and the cloud servers sometimes become lazy because of the computation that affects correctness of final result. Therefore, the computation has to be made transparent. Healthcare data mainly contain of large media files such as X-ray, CT scans, radiology, and other type of images and videos. Such files are called as the Electronic Health Records that are stored in distributed storage. Possibly, this patient perception is fueled by the fact that healthcare data may be disclosure to unauthorized person [4].

In order to secure the patient's sensitive data from unauthorized access, an appropriate encryption standard must be applied to data stored in cloud. This sensitive information is most confidential and needs to be protected. To put everything in the cloud in an unencrypted is a big risk. Over the past four decades, a lot of efforts have been put into developing healthcare information security systems. There is a great variety of commercially available programs to assist clinicians with diagnosis, decision making, pattern recognition, medical reasoning, filtering, and so on for general and very specialized domain applications. If a healthcare system is not secured, an adversary could read, modify, and inject messages into the network. Such incorrect information, even when not for nefarious reasons, can lead to serious consequences for patients and for safe services such as remote healthcare monitoring due to using heterogeneous devices that use a variety of communication rules. Most of the rules that are designed for cloud-based communication cannot be directly applied in the cloud-based healthcare network. In cloud-based healthcare system, remote nodes have limited computation, processing, and communication rights [5].

The existing techniques for healthcare data include pseudo copulation (replacing the most identifying fields in a data) and encryption (encoding the data in such a way that only authorized remote institutions can access it). The existing safeguards are referred to as medical confidentiality or doctor-patient privilege, which prohibit a medical professional to disclose information about a patient's case. This is an important obligation within the medical professional in order to create trust between a doctor and his patient and a trusting environment in which the patient feels comfortable. If a patient cannot trust a physician's discretion, he will not seek

medical care altogether or will withhold information during a consultation. If a physician does not have accurate information about a patient's health, this may lead to an inaccurate diagnosis and improper treatment, which may lead to great harm to the patient's health [6].

Figure 1 shows a typical information flows in the healthcare network. Patient information serves as a range of purposes apart from diagnosis and treatment provision. Patient information could be used to improve efficiency within the healthcare system. Patient information could be shared with finance facilitators to justify payment of service rendered. Health service providers may share health information through improved service quality. Furthermore, these providers may share health information through Regional Services to facilitate care services in the regional areas [7].

Credentialing is a vital process for all healthcare systems that must be performed to ensure that those healthcare workers who will be providing the clinical services are qualified to do so. The cloud-based healthcare system is capable to ensure patient safety and deliver an acceptable standard of care. While employing excellent medical staff is vital for success, the healthcare system must have to define the required minimum credentialing and privileging requirements to validate the competency of healthcare providers. In the classical systems, only hospitals used to perform credentialing, but our proposed system has capability to provide all healthcare facilities and also to perform credentialing [8].

In this framework, we classify different modules based on the probability (i.e., trust level) of each provider in violating the patient's data in detail. Honestly, I cannot understand exactly what this statement means. Remote nodes (healthcare physicians, nurses, family members, and other authorized individuals) are different from main modules (patients, health service providers, finance facilitators, regional services, and evaluative decisions), and so it is necessary to make clear remote nodes and modules because the patient Bayesian model only evaluates the trusty of

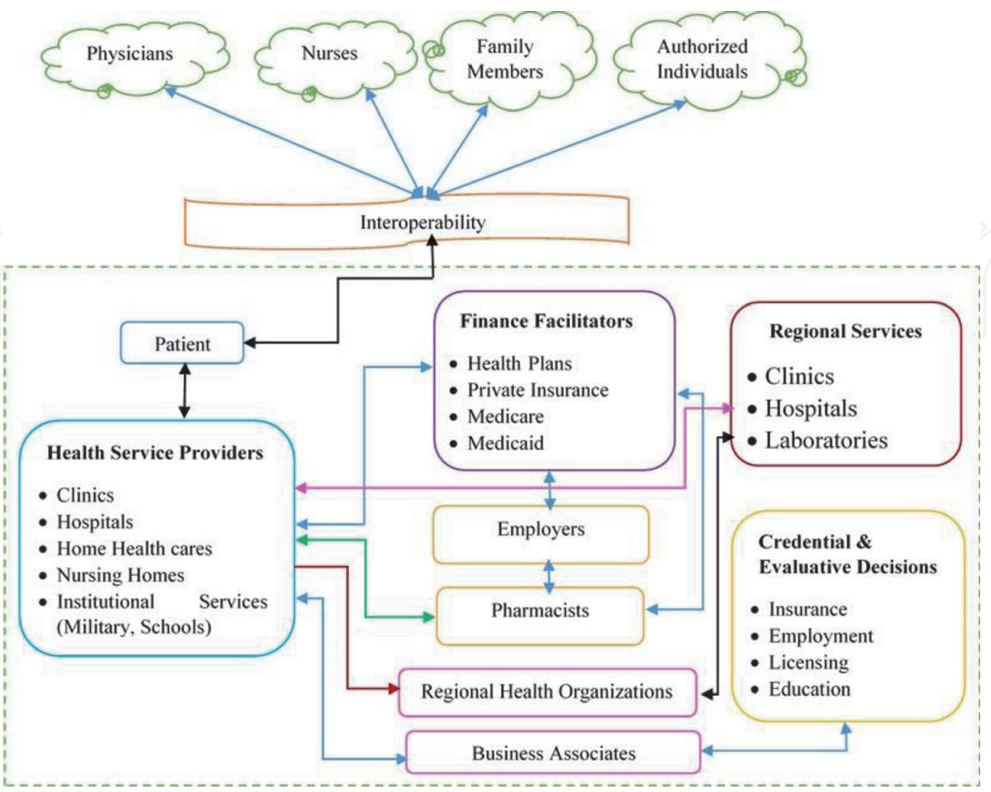


Figure 1.
Cloud-based healthcare system.

remote nodes and whole network based on service level expectation (SLE) as evidence, following the statement "we take advantage of the nature of the Bayesian inference to calculate the probability of wireless communication between the healthcare system and its remote institutions" [9].

3. Problem statement

Healthcare system has become the inspiration for patients' data in terms of wirelessly communication for decision making and logical functionality of the remote institutions such as health physicians, nurses, family members, and authorized individuals. Conventional healthcare systems are using various encryption methods to secure patients' data. Observing the limitations of the existing encryption methods, we take advantage of the nature of the Bayesian inference to calculate the probability of wireless communication between the healthcare system and its remote institutions. The dynamics of the cloud environment requires the healthcare system being able to self-adapt, being aware of its surrounding environment's changing parameters, and being able to create new rules based on past experience. To eliminate the problem of repetition in the cloud environment, the security algorithm must maintain the remote institution limitations and at the same time must provide high level of data protection. Constraint-based adaptive boost algorithm has progressed to an advanced level data analysis for cloud-based healthcare system. The implementation of patient Bayesian Inference for cloud-based healthcare system will be suitable to demonstrate its performance in the real-world patients' datasets. Protection of patient's sensitive data is one of the main obstacles to the growth of cloud computing in the health field because of the need for high level of data integration, interoperability, and sharing among healthcare institutions. It is necessary to create standard guidelines and identify security challenges for improving information security in healthcare system. There are multiple remote institutions (nodes) that have to deal with healthcare data such as healthcare physicians, nurses, family members, and other authorized individuals. Similarly, within healthcare system, there are multiple entities that have to deal with healthcare data such as healthcare providers, hospital administration staff, finance providers, and patients themselves. Cloud services suffer from certain vulnerabilities [10]. By contrast, Bayesian model as an uncertain reasoning tool is more efficient for dynamic trust evaluation. Bayesian inference combined with cloud model and Bayesian network is proposed in this research.

4. Patient Bayesian inference

In cloud-based healthcare systems, patients' electronic data have been widely adapted to improve the quality of patient care and increase the productivity and efficiency of healthcare delivery. In cloud-based systems, patients' data can be helpful to resolve many of the existing problems associated with disease diagnosis and also maintaining the privacy and sensitivity of the patient's medical information. PBI can be beneficial in the healthcare system for tracking fatigue by using multiarmed bandits, which facilitate the healthcare doctors in treatment by dynamically taking more samples from those treatments, which are most likely to be the best. PBI may facilitate the doctors in better understanding the patient's data and make decisions based on it. Because of security in cloud computing, outcomes can be measured in real time, rather than waiting for enough data. Recently, health data privacy has become an important issue in the cloud-based healthcare systems.

As a result, data mining techniques include swapping attribute values and principal component analysis-based techniques, adding random components have gained much more attention in the healthcare data analysis [11].

In healthcare system, PBI is an extremely powerful set of tools that use some knowledge or beliefs to calculate the probability of biomedical and healthcare events, statics, and Service Level Expectation (SLE). PBI can be used for mapping our understanding of a problem and evaluating observed data into a quantitative measure of how certain we are of a particular fact in probabilistic terms, where the probability of a proposition simply represents a degree of belief in the trust of that proposition. PBI can also be used as data mining technique for analyzing network healthcare system variables, virtual assistants, and other variable analytics [12]. PBI uses data and evidence that certain facts are more likely than others. Prior distribution reflects our belief before seeing any data, whereas posterior distributions reflect our belief after we have considered all the evidence.

Cloud-based PBI consists of five main modules: (i) patients; (ii) health service providers; (iii) finance facilitators; (iv) regional services; and (v) traditional and evaluative decisions and four submodules: (a) employers; (b) pharmacists; (c) regional health organizations; and (d) business associates. In this framework, we classify different modules based on the probability (i.e., trust level) of each provider in violating the patient's data. Bayesian rules allow calculating the posterior probability of any information violation events as hypothesis (H) based on a set of historical data (D).

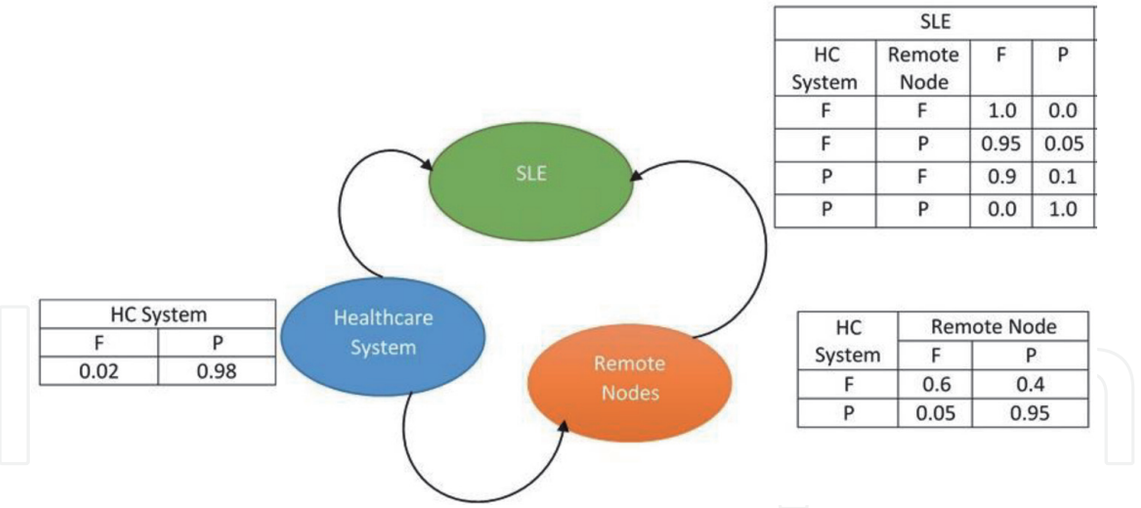
$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} \tag{1}$$

where $P(H|D)$ is the posterior probability of H given knowledge data D; $P(H)$ is the prior probability for H; $P(D|H)$ is the likelihood probability of H given D; and $P(D)$ is the marginal probability that would have happened whether or not H is true. In cloud-based healthcare system, we use Bayes' rule to find the probability function as in Eq. (2):

$$P(\text{SLE}, \text{healthcare system}, \text{remote nodes}) = P(\text{SLE} \mid \text{healthcare system}, \text{remote nodes}) * P(\text{remote nodes} \mid \text{healthcare system}) * P(\text{healthcare system}) \tag{2}$$

SLE is abbreviation of service level expectations. In cloud-based healthcare system, SLE is responsible to provide the quality of services to the remote nodes. It can also be variable that has enough relevance for the service and can be quantitatively and objectively measured. It strengthens the processes to improve the outcomes. In Bayesian Inference, our initial beliefs are represented by the prior distribution P(healthcare system) as shown in **Figure 2**.

In **Figure 2**, remote nodes and healthcare network are hidden variables, and the only observable variable is the SLE metric. An SLE node forecasts how long it should take a share healthcare information to the remote nodes. The SLE itself has two parts: a period of elapsed time and a probability associated with that period (e.g., 38% of healthcare information is shared in 5 min or less, which can also be stated as "5 min with 38% confidence/probability"). However, the healthcare network is a complete system for the variables and their dependencies. Healthcare system can also calculate the services provides to the services provided to the remote nodes like "what is the probability that network successfully passes and the given SLE has failed, P(healthcare system = true | SLE = false), which shows that the sharing of the healthcare information with the remote nodes is not completed within the threshold level. In general, the ultimate purpose of the proposed patient



SLE			
HC System	Remote Node	F	P
F	F	1.0	0.0
F	P	0.95	0.05
P	F	0.9	0.1
P	P	0.0	1.0

HC System	
F	P
0.02	0.98

HC System	Remote Node	
	F	P
F	0.6	0.4
P	0.05	0.95

Figure 2.
Communication between healthcare system and remote nodes.

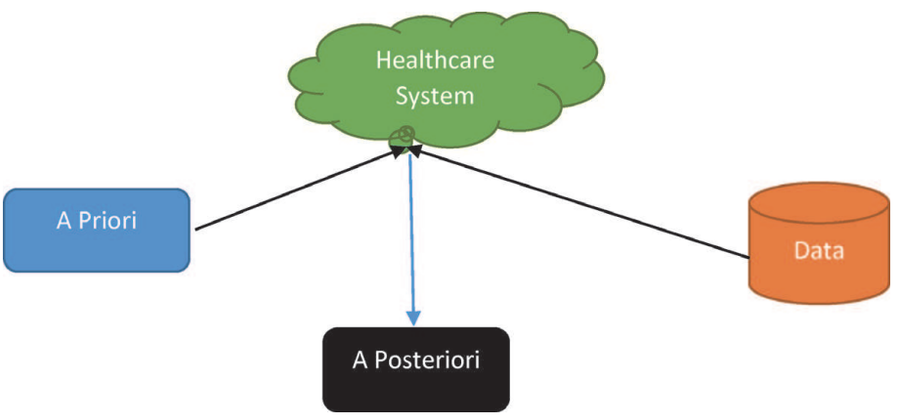


Figure 3.
Bayesian rules.

Bayesian model is to calculate the posterior (conditional) probability of the healthcare system given SLE, P(healthcare system| SLE), which reflects the trusty of the healthcare system. Eq. (3) calculates the posterior probability P(healthcare system | SLE), according to Bayes' rule (**Figure 3**).

It is necessary to choose a probabilistic model represented by Eq. (2) that relates to the random variables and the model parameters associated with it. At the end, Bayes' rules are applied to combine the prior knowledge and the new observed data to find the posterior probability distribution, following Eq. (3).

$$P(\text{healthcare system} \mid \text{SLE}) = \text{Sum of } P(\text{SLE}, \text{healthcare system}, \text{remote nodes}) \text{ over all values of remote nodes divided by sum of } P(\text{SLE}, \text{healthcare system}, \text{remote nodes}) \text{ over all values of remote nodes and SLE.} \tag{3}$$

5. Constraint-based adaptive boost algorithm

The constraint-based adaptive boost (CBAB) algorithm is a simple, flexible, and effective classifier [13]. In cloud-based healthcare system, CBAB is used for patient's data analysis. In healthcare system, each patient has different set of records with some common features and unique attributes such as name, age, disease, etc.

Let $D^1_{n(1)}, D^2_{n(2)}, \ldots D^M_{n(M)}$ are the datasets of M patients and the dataset of P^{th} node contains a total of $n^{(p)}$ samples, and it can be represented as:

$$D^P_{n(p)} = \left\{ \left(X^p_{1(p)}, Y^p_{1(p)}\right), \left(X^p_{2(p)}, Y^p_{2(p)}\right), \ldots \left(X^p_{n(p)}, Y^p_{n(p}\right) \right\} \tag{4}$$

where X^p_i is the patient's data at P^{th} node and Y^p_i is the decision making that is being consider here. The CBAB algorithm is applied to analyze the health information of each patient for "t" boosting iterations. In the decision making, each unidentified data is represented by $\left(f_n, \theta, \delta \right)$, where f_n represents the selected health parameter, θ is the decision threshold, and δ is the sign of decision, i.e., +1 or −1. CBAB calls a given learning algorithm in a series of loops t = 1, 2, … t. For any health information X_i, the hypothesis $\mathrm{h}(X_i)$ means the decision is either +1 or −1. For the P^{th} patient, $H^P(.)$ is the set of T unidentified data:

$$\left\{ h^{P(1)}(.) \propto^{P(1)}, h^{P(2)}(.) \propto^{P(2)}, \ldots h^{P(T)}(.) \propto^{P(T)} \right\} \tag{5}$$

where $h^{P(t)}$ is the unidentified data at t^{th} iteration and $\propto^{P(t)}$ is the corresponding weight of the unidentified data. For a particular patient's information X_i, the prediction made by the P^{th} patient can be defined as:

$$H^P(X) = sign\left\{ H^P(X) \right\} = sign\left\{ \sum_{t=0}^{T} \propto^{P(t)} h^{P(t)}(X_i) \right\} \tag{6}$$

In a cloud-based healthcare system, all the nodes can share a patient's data to each other, and hence each node will receive M-1 information from other nodes. Therefore, each node would integrate specific information. In this way, the healthcare system would value the sensitivity of the patient's information for decision making. To analyze the original patient's data among different nodes in healthcare system is infeasible due to patient's privacy, therefore, we alternate to applying all the other nodes in the training set of P_{th} node, and compare the error rate of each node with the training rate of P_{th} node as shown in Eq.(7):

The node receiving information from any other node might be changed data, hence before using such data, the P^{th} node should select a suitable subset of relevant data based on f_n. For the P^{th} node, the error rate of q^{th} node is given by:

$$\in^{(q)}_P = \frac{1}{n^{(P)}} \left[\sum_{i=1}^{n^{(P)}} I(sign\left\{ H^q\left(X^P_i\right) \neq Y^P_i \right\} \right] \tag{7}$$

where $H^q(.)$ is the selected information patterns from patient's shared data by node q, and I(.) is the indicator function. The training rate of the P^{th} trained node is given by:

$$\in_P = \frac{1}{n^P} \left[\sum_{i=1}^{n^{(P)}} I(sign\left\{ H^P\left(X^P_i\right) \neq Y^P_i \right\} \right] \tag{8}$$

For every node, we compute the difference between $\in^{(q)}_P$ and $\in_P$. If $\left(\in^{(q)}_P - \in_P \right)$ is less than a certain threshold level, then we can assume that the patient's data shared between P^{th} and q^{th} nodes are similar and we can use q^{th} node as trust node for P^{th} node

6. Conclusion

In our research, we have proposed a patient Bayesian Interference for analyzing the healthcare system. The Bayesian Inference is used to model the uncertainties that come with the problems and dealing with missing data and also allow integrating data from remote resources. We have also used the concept of constraint-based adaptive boosting to demonstrate the patient's Bayesian inference performance in the real datasets from healthcare system to remote resources. In the future, we will try to find more accurate ways to protect the patient's data more accurately without compromising on patient's privacy.

Author details

Shahid Naseem
Department of Information Sciences, Division of Science and Technology, University of Education, Lahore, Pakistan

*Address all correspondence to: shahid.naseem@ue.edu.pk

References

[1] Karim A, Abderrahim H, Hayat K. Big healthcare data: Preserving security and privacy. Journal of Big Data. 2018;**5** (1):1-8

[2] Jawwad A, Ali K. Understanding privacy violations in healthcare big data systems. IEEE. 2018;**20**(3): 273-281

[3] O'Mareen S, Richie O, Peter D. Patient consent to publication and data sharing in industry and NIH-funded clinical trials. Springer Nature. 2018; **269**(19):10-25

[4] Hina A, Jawad H, Junaid C, Kashif S, Mehmet A, Jalal M, et al. Risk analysis of cloud sourcing in healthcare and public health industry. IEEE Access.2018; **6**:35-55

[5] Isra's Ahmed S, Ahmed Mousa A. Security and privacy issues in Ehealthcare systems: Towards trusted security. IJACSA. 2016;**7**(9):229-236

[6] Aaron B, William H, Michal M, Abdemour K, Thar B, Casimino A. An investigation into healthcare data patterns. MDPI. 2019;**11**(2):1-23

[7] Faruk A, Aftab A, Haider A, Nur Al Hassan H. A cloud-based healthcare framework for security & patient's data privacy using Wireless Body Area Network. Elsevier. 2014;**34**:511-517

[8] Roshan P, Sandeep S. Credentialing. India: NCBI; 2019

[9] Ella G, Jae S, Amanda D, Wenday S. Averse health events associated with clinical placement: a systematic review. Elsevier: Nurse Education Today. 2019; **76**(1):178-190

[10] Tatiana E, Geboren M. Security and Acceptance of Cloud Computing in Healthcare. Berlin: Der Technischen Universiat; 2015

[11] Alther M, Redday C. Clinical decision support systems. In: Redday C, Aggarwal C, editors. Healthcare Data Analytics. London: Chapman and Hall Press; 2015. pp. 225-260

[12] Rafiqullah S, Sagar R, Anshul S, Nasar Uddin A. Bayesian method for modeling male breast cancer survival data. APJCP. 2014;**15**(2):663-669

[13] Li Y, Bai C, Reddy C. A distributed ensemble approach for mining healthcare data under privacy constraints. Information Sciences. 2016; **330**:245-259