SURVEY ON SECURE MACHINE LEARNING CLASSIFICATION ALGORITHMS FOR CLOUD-BASED REMOTE HEALTHCARE SERVICES

Sonali Shinde, Dr. Satish N. Gujar

Department of Computer Engineering, TSSM's BSCOER, Narhe, Pune

ABSTRACT

as more and more people use healthcare data analytics, the method of classifying data has become very important. A data classifier is usually made and then sent to a cloud. Then, a service provider can easily offer a wide range of services and deal with a lot of classification requests from users. With privacy and security issues, the valuable classifier and the sensitive user data can't be directly sent to the cloud. In this paper, we focus on the Naive Bayes (NB), which is one of the most popular classifiers. We show how to outsource NB classification in the cloud in a way that is both efficient and private. Specifically, the service provider can change the traditional NB classifier into fixed hyper-rectangles, and the order-preserving encryption is used to encrypt these hyper rectangles as the encrypted classifier, so the service provider can use the encrypted classifier. After that, the encrypted classifier is sent to the cloud, and users can send encrypted queries to the cloud and get the classification results back.

Keywords*– Cloud-based online diagnosis services, naïve bayes classification, encryption.*

INTRODUCTION

Data classification is one of the most important big data analytic techniques. It can be used to make accurate predictions for unknown data samples, which can help a wide range of applications, such as online disease diagnosis and vehicular intelligence. Many different types of data classification functions (also known as classifiers), which are owned by a service provider, can be sent over to a third-party cloud server for easy configuration and management. At the same time, users will get better service because the cloud has powerful computing and storage abilities,naive bayes (NB) is one of the most common classifiers, especially for online diagnosis services. It is used a lot. When a hospital outsources its NB diagnosis classifier to the cloud, it can provide online diagnosis services, and patients can send their medical information to the cloud and get a diagnosis prediction afterward.

Even though the cloud-based data classification model has many advantages, the privacy issue is a big problem for it. Cloud-based model means that users have to give their data to the cloud in order to complete the classification process from afar, but they should not give or leak their sensitive data, like their medical history, to any other cloud. This is because public clouds, like Amazon Cloud and

Microsoft Azure, have a good chance of being hacked by the adversary, as shown in the recent news. Service providers don't want to show off their trained NB classifiers because they need a lot of training samples to do so. They'd rather keep it a secret than show it off. Two privacy requirements must be met in this case to keep the privacy of the users' information, such as the data features (input) and predictions (output).

In this paper, we show how to outsource NB classification in the cloud in a way that protects your privacy. Based on an efficient cryptographic algorithm, the traditional NB classification problem can be changed into a new range query problem. That is, the classifier can be shown as a group of hyper-rectangles made by vectors, and the boundary of these hyper-rectangles can be used to classify data. Scheme: Encryption is used to hide the process of transforming as well as the range query. This way, both the service provider and the user can keep their private information safe.

LITERATURE REVIEW

This paper introduce most popular classifiers, the Support Vector Machine (SVM) is one of the most efficient and privacy-preserving ways to outsource SVM classification in public cloud computing environments. Additionally, the SVM classifier is transformed into fixed hyper-rectangles, which are then encrypted using the order-preserving encryption method to create the encrypted classifier. It is then outsourced to a public cloud where a user can submit an encrypted range query and get the categorization results back. Security research and thorough experimentation show that our system protects the security of classifier and user data and achieves efficient SVM classification in terms of computational cost [1].

Support vector machine (SVM) learning, a potent classification approach for cancer genomic classification or subtyping, uses machine learning with maximising (support) of separation margin (vector). The classification characteristic of SVMs is now being used more widely in cancer genomics, resulting in the discovery of new biomarkers, new therapeutic targets, and a better understanding of cancer driver genes as a result of advances in high-throughput genomic and epigenomic data collection. SVMs have recently made significant strides in cancer genomics research, which authors have summarised in this paper[2].

Cloud-assisted online diagnosis services are described in this research. The suggested work-flow includes an efficient and secure decision-tree categorization mechanism. Specifically, the medical institution uses searchable symmetric encryption to encrypt a locally pre-trained decision tree classifier into a decision table. Encrypted data is then sent to a cloud server, where a user can enter encrypted physiological features and receive an encrypted diagnosis prediction. In order to show the security of our decision tree classifier and user data, authors give rigorous security proofs. This method outperforms linear classification in terms of speed, according to a performance evaluation[3].

It is in this study that authors present PPDP, a privacy-preserving and efficient method of disease prediction. It is possible to train prediction models by employing Single-Layer Perceptrons in a privacy-preserving manner by encrypting patients' medical records and sending them to the cloud server. Models can be used to estimate the likelihood of diseases in new medical data. Medical data encryption, illness learning, and disease prediction algorithms based on random matrices are some of the innovations in PPDP. PPDP provides the necessary level of privacy protection, according to a security assessment [4].

EMRs from actual electronic health (eHealth) systems were studied in this paper, and authors found that (1) multiple patients would generate a large number of duplicate EMRs, and (2) cross-patient duplicate EMRs would be generated numerously only in the case that the patients consult doctors within a department at the same hospital. For cloud-assisted eHealth systems, authors then present the first efficient and secure encrypted EMRs deduplication technique (HealthDep). Cloud server storage costs can be reduced by more than 65 percent by integrating HealthDep's analytic results into the EMRs deduplication process, and the confidentiality of EMRs can be maintained [5].

Machine learning models are becoming increasingly important in today's data-driven economy. Customers' behaviour can be predicted with extreme precision using the massive volumes of data that businesses collect. To prevent competitors from copying their services, it is essential to protect the data needed to develop these models. The privacy of the data that needs to be categorised must also be protected as this type of approach enters more sensitive industries, such as the medical industry. Homomorphic evaluation of Support Vector Machine (SVM) models is used to address this issue, ensuring the client and service provider remain in the dark regarding the model's workings save for the classification of their data. A new problem has emerged with the usage of SVMs in Fully Homomorphic Encryption (FHE), which previously concentrated on either bit-wise or value-wise computations. SVMs integrate the two, requiring the evaluation of both the kernel function and real arithmetic before extracting the sign bit [6].

Nonlinear kernel support vector machines (SVMs) are used in this study to develop an online medical pre-diagnosis system, called eDiag, that is both fast and secure (SVM). Online prediagnosis services can be conducted securely with eDiag, which protects the confidentiality of sensitive personal health information. An efficient and privacy-preserving classification strategy is developed using lightweight multi-party random masking and polynomial aggregation approaches based on an improved expression for the nonlinear SVM. To perform the diagnosis, the service provider decrypts the encrypted user query, and the user must then decode the diagnosis result [7].

EXISTING SYSTEM:-

Lot of work has been done in this field because of its extensive usage and applications. In this section, some of the approaches which have been implemented to achieve the same purpose are mentioned. These works are majorly differentiated by the techniques for Cloud Based Remote Healthcare Services classification systems.

- Health care industry today generates large amounts of complex data about patients, hospital resources, disease diagnosis, electronic patient records, medical devices etc.

- With the widespread use of hospital information system, there is a huge amount of generated data which can be used to improve health care services, thus developing data mining applications to provide people more customized health care service.

PROPOSED SYSTEM:-

In this paper, we show how to outsource NB classification in the cloud in a way that protects your privacy. Based on an efficient cryptographic algorithm, the traditional NB classification problem can be changed into a new range query problem. That is, the classifier can be shown as a group of hyper-rectangles made by vectors, and the boundary of these hyper-rectangles can be used to classify data. Scheme: Encryption is used to hide the process of transforming as well as the range query. This way, both the service provider and the user can keep their private information safe.

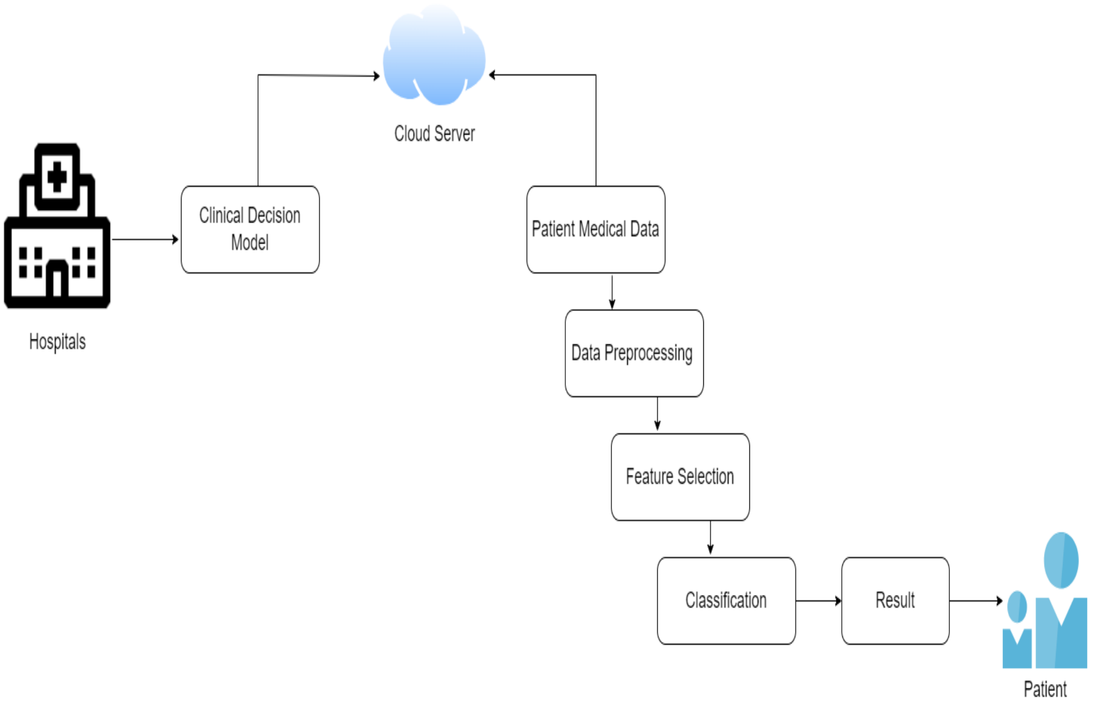

Figure 1. System Architecture

CONCLUSION

In this paper, we have come up with a way to keep machine learning classification schemes safe in the clouds. In the proposed scheme, the secure outsourced naive bayes classification problem is changed to a secure range query problem. Then, order-preserving encryptions are used to keep data private. Both security and performance tests show that our scheme is very efficient while protecting both the data and the classifiers. Also, experiments show that our scheme is very efficient in terms of computational overhead. It only takes a few microseconds to encrypt an NB classifier.

FUTURE WORK

For future work, we'll do more experiments to see how well our scheme works and how well other methods work. We'll also design an efficient NB classifier with a better security guarantee.

REFERENCES

[1] J. Liang, Z. Qin, J. Ni, X. Lin, and X. Shen, “Efficient and privacy preserving outsourced SVM classification in public cloud,” in Proc. of IEEE ICC, 2019, pp. 1–6.

[2] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, “Applications of support vector machine (svm) learning in cancer genomics,” Cancer Genomics-Proteomics, vol. 15, no. 1, pp. 41–51, 2018.

[3] J. Liang, Z. Qin, S. Xiao, L. Ou, and X. Lin, “Efficient and secure decision tree classification for cloud-assisted online diagnosis services,” IEEE Trans. on Dependable and Secure Computing, pp. 1–13, accepted 2019, to appear, DOI: 10.1109/TDSC.2019.2922958.

[4] C. Zhang, L. Zhu, C. Xu, and R. Lu, “PPDP: an efficient and privacy-preserving disease prediction scheme in cloud-based ehealthcare system,” Future Generation Comp. Syst., vol. 79, pp. 16– 25, 2018.

[5] Y. Zhang, C. Xu, H. Li, K. Yang, J. Zhou, and X. Lin, “Healthdep: An efficient and secure deduplication scheme for cloud assisted ehealth systems,” IEEE Trans. on Industrial Informatics, vol. 14, no. 9, pp. 4101–4112, Sept 2018.

[6] J.-C. Bajard, P. Martins, L. Sousa, and V. Zucca, “Improving the efficiency of svm classification with fhe,” IEEE Trans. on Information Forensics and Security, pp. 1–14, accepted 2019, to appear, DOI: 10.1109/TIFS.2019.2946097.

[7] H. Zhu, X. Liu, R. Lu, and H. Li, “Efficient and privacy-preserving online medical prediagnosis framework using nonlinear SVM,” IEEE J. Biomedical and Health Informatics, vol. 21, no. 3, pp. 838–850, 2016.

[8] J. Liang, Z. Qin, S. Xiao, J. Zhang, H. Yin, and K. Li, “Privacypreserving range query over multi-source electronic health records in public clouds,” Journal of Parallel and Distributed Computing, vol. 135, pp. 127–139, 2020.

[9] N. G. Tsoutsos and M. Maniatakos, “Efficient detection for malicious and random errors in additive encrypted computation,” IEEE Trans. on Computers, vol. 67, no. 1, pp. 16–31, 2018.

[10] X. Li, Y. Zhu, J. Wang, Z. Liu, Y. Liu, and M. Zhang, “On the soundness and security of privacy-preserving svm for outsourcing data classification,” IEEE Transactions on Dependable and Secure Computing, vol. 15, no. 5, pp. 906–912, 2018.