A Privacy-Preserving High-Order Neuro-Fuzzy C-Means Algorithm with Cloud Computing

Dr. Kotrappa Sirbi[1] Mr. Abhijit J. Patankar[2] Mrs. Archana J. Jadhav[3] Mr.Rahul Jadhav[4]
[1]Professor [2]Research Scholar [3]Assistant Professor [4]M.E.Student
[1,2,3,4]Department of Computer Science & Engineering
[1]KLE Dr. M.S. Sheshgiri College of Engineering & Technology, Belagavi, India
[2]VTU, Belgaum, Karnataka, India [3,4]Alard College of Engineering, Pune, India

Abstract— In the real world massive heterogeneous details and data are generated from the cloud. Heterogeneous data is basically a different type of data combined together. Processing on the heterogeneous data is done with the help of neuro-fuzzy technology. This becomes a hot topic for cloud. We propose a privacy-preserving high-order neuro-fuzzy c-means algorithm for clustering heterogeneous data on the cloud. Privacy-preserving high-order neuro-fuzzy c-means algorithm on cloud computing clusters the heterogeneous data set by representing each heterogeneous data object as a tensor and uses the tensor distance to capture the correlations in the high-order tensor space. Furthermore, the cloud computing is employed to improve the clustering efficiency for massive heterogeneous data from cloud. The BGV encryption mechanism or technique is used to protect the private data when performing the high-order neuro-fuzzy c-means algorithm on to the cloud. We propose a practical privacy-preserving c-means clustering scheme that can be efficiently outsourced to cloud servers. Our scheme allows cloud servers to perform clustering directly over encrypted datasets, while achieving com-parable computational complexity and accuracy compared with clustering's over unencrypted ones. We also find out the secure integration of MapReduce into our mechanism, which makes our mechanism or we can say scheme mostly suitable for cloud computing environment. By security survey and numerical survey carry out the performance of our mechanism in terms of security and efficiency.

Key words: Privacy Preserving, Fuzzy Systems, Cloud Computing, C-Means

I. INTRODUCTION

The explosive growth of sensor technologies has promoted the development of Internet of Things (IoT). Aim of IoT is to better life quality by combining the computing theory and the networking technology. IoT used in many applications such as industrial control, intelligent transportation and smart medical. Nowadays, a large number of heterogeneous data, often referring to big data, is generating from IoT, which requires novel models and technologies to process, especially fuzzy and neuro computing, for the further promotion the design and applications of IoT. However, the heterogeneous data is usually very complex

As a fundamental technique of data mining, clustering groups the objects into several categories such that the objects in the same category share as much similarity as possible. There are so many algorithms have been proposed for heterogeneous data clustering.

In the previous work, we proposed a high-order possibility c-means algorithm by extending the conventional possibilistic c-means algorithm from the vector space to the tensor space for multimedia heterogeneous data clustering. Furthermore, we employed cloud computing to improve the clustering efficiency for massive heterogeneous data. To protect the private data during clustering on cloud, we proposed a privacy-preserving high- order possibility c-means algorithm (PPHOPCM) by using the fully homomorphism encryption scheme to encrypt the original data. PPHOPCM uses the Taylor theorem to approximate the membership matrix updating function as a polynomial function for supporting the secure computation of the membership matrix updating function. Some experimental results demonstrated that PPHOPCM achieves state-of-the-art performance for clustering multimedia heterogeneous data. However, PPHOPCM always produces a coincident result when clustering massive heterogeneous data in IoT. Aiming at this problem, the paper presents a high-order fuzzy a high-order fuzzy c-means (HOPCM) algorithm to cluster heterogeneous data in IoT. We extend the fuzzy c-means algorithm, instead of the possibilistic c-means algorithm, from the vector space to the high-order tensor space. In the high-order fuzzy c-means, we use the tensor distance to measure the distance between each pair of objects for avoiding the coincident clustering result. Furthermore, we propose a privacy-preserving high-order fuzzy c-means algorithm (PPHOFCM) to improve the clustering efficiency of the high-order fuzzy c-means algorithm using the cloud computing techniques without disclosure of private data. To achieve the goal of preserving the private data, we also use the currently most efficient fully holomorphic encryption scheme, BGV in the scheme. The member-ship matrix updating function of PCM is a function of one variable while that of FCM is a function of many variables. Therefore, different from the PPHOPCM which uses the Taylor theorem to sup-port the secure computation of membership matrix updating function, PPHOFCM uses the Multiple Taylor theorem to approximate the membership matrix updating function as a polynomial function. The privacy- preserving high-order neuro-fuzzy c-means algorithm brings a lot of issues and challenges, especially for clustering analytic of massive heterogeneous data collected from IoT by employing the powerful computing and store of the cloud. We discuss the key challenges in the following three aspects. (1) To cluster the heterogeneous data, it requires the ex-tension of the fuzzy c-means from the vector space into the tensor space. (2) To protect the sensitive data and intermediate results, it requires secure computation of various operations needed by the high-order neuro-fuzzy c-means algorithm, including secure additions, secure multiplications, and the secure nonlinear function. (3) To improve the efficiency of high-order neuro-fuzzy c-means clustering for massive heterogeneous data, it requires choosing the efficient full

holomorphic encryption scheme according to the major operations of the algorithms in the privacy-preserving high-order neuro-fuzzy c-means algorithm.

There are two type of dataset i.e., hIoT and mmIoT, to evaluate the performance of the introduced schemes by comparison with SHDC and HOPCM. Results demonstrate that the proposed high-order fuzzy c-means scheme performs better than the conventional FCM in terms of clustering accuracy. Furthermore, privacy-preserving high-order fuzzy c-means can enhance the efficiency of FCM by uploading the computationally intensive operations to the cloud without the leak of the sensitive information. More important, the performance of our method can be further enhanced by employing more cloud servers.

Document clustering has been used in most of the different areas of text mining and information recovery. Originally it was used for enhance the precision and recall in information recovery systems and detecting nearest neighbors of a document. After this it will also been used for arranged the results returned by a search engine and produced hierarchical clusters of documents.

We have also try it out a completely different approach by first clustering the words of the documents by using a standard clustering approach and thus reducing the noise and after that using this word cluster to cluster the documents. We found that this approach also give better results than the classical K-Means and Agglomerative Hierarchical clustering methods.

Our research work and this paper focuses on how we want to implement data optimization using indexing and optimization technique. Also further we will be classifying this data using c-means classification algorithm over the heterogeneous data. We will be focusing on generalization and suppression methods to clustering this data which will produce more effective noise free results.

II. LITERATURE REVIEW

Widely used algorithm in the areas of the statistics, data mining and pattern reorganization is the fuzzy C-means. Its presentation is not satisfactory when used with multi-dimensional and bulky dataset. In this paper proposes a scenario to beat the above issue of Fuzzy-C-Means algorithm called as sparse c-means fuzzy algorithm which is overall depends on the framework uses the sparse clustering.

This methodology combines features into the traditional fuzzy algorithm with the help of the weighting of sparse, it makes easier to understand. Based on the analysis we can conclude that developed approach can get key features and increase the usefulness at the time of clustering on massive dataset. The FCM is most commonly used method for clustering which is firstly developed by the Dunn for the pattern recognition and data mining.

The traditional methods cannot handle multidimensional data efficiently. High dimensional data contains the information which is redundant in which only few numbers of features are useful. Common agenda is to differentiate the important features.

The sparse clustering can handle the problems related with the high dimensional data. Which can able to embed the features of traditional clustering methods such as the k-means and hierarchical clustering using this framework we can obtain greater efficiency?

Gene expression analysis which is the application of the large scale and high dimensional clustering. In this paper to analyses the efficiency of the SFCM used two different gene expression profiles datasets. Which one is the peripheral blood mononuclear cell (PBMC) another is Breast Cancer. Traditional dimensionality approach use to combine the features to produce the new features such as principle component analysis and PCA-type methods. Which is used to transform features into lower dimension?

The developed approach is based on the sparse clustering framework which can able to handle the problems occurs during the operation on the high dimensional data. It can efficiently cluster the high dimensional data. It has two main advantages first one it has ability to maintain FCM with fuzzy behavior, second one with the help of sparse matrix weighting method we can access the features to build the model. Result analysis is based on the artificial dataset and the gene expression dataset.

In terms of the web mining and the information retrieval important aspect is the co-clustering based on heterogeneous data need have to work with high order co-clustering.

In this paper represented a high order co-clustering approach. Which is able to cluster the available data into different forms? In this use high order tensor to model the high order relationship, each element represents the relationship between the collected set of data objects.

With the help of the high order relationship we can combine the data objects whose types are different. At the final clustering is performed on the lower dimensional data using the k -means approach. Experimental analysis is performed on both data using proposed method.

In this approach higher order relationship of multiple type of data objects is structured. The similar data objects which analyses from the different types using unified modeled without using pair-wise fusion which might not dependable with the original data structure.

With the help of higher order modeling Clique Expansion technique is used to approx. the tensor by matrix which is not analyses by spectral cut on graph. This methodology can be used in future to work on the multi-type heterogeneous data which is common in real-time.

In recent days the share marketing values can't be anticipated accurately. This prompts overwhelming misfortune for the customer. In this undertaking subsequent to breaking down tremendous measure of information about advertising field the master will foresee the more precise esteem for the share because of these inconveniences, another packing estimation called C-Means algorithm is proposed. We design a scattered HOPCM procedure in perspective of map Reduce for a considerable measure of heterogeneous data finally; we devise a security sparing HOPCM computation (PPHOPCM) to guarantee the private data on cloud by applying the BGV encryption to HOPCM. Using this algorithm the share marketing expert can cluster the churn data. By this process client are saved from more losses(Risk) with the help of share marketing expert's prediction .Now the client can buy the share without any risk with the help of share marketing expert's suggestion.

The release and clarity/transparency of information flow on the Web has increase concerns of privacy. Given a set of data items, clustering algorithms do group of similar items together. Clustering has many applications, such as customer behavior survey, targeted marketing, forensics, and bioinformatics. This approached we present the design and examine of a privacy-preserving k-means clustering algorithm formally called as KNN, where only the cluster means at the various steps of the algorithm are disclose to the join parties. The critical step in our privacy-preserving k-means is privacy-preserving computation of cluster means. We present two set of rules first one based on oblivious polynomial evaluation and the second one is based on homomorphic encryption for privacy-preserving computation of cluster means. We have a JAVA implementation of our algorithm. Using our implementation, we performed a thorough evaluation of our privacy preserving clustering algorithm on three datasets. Our judgment demonstrates that privacy preserving clustering is feasible, i.e., our holomorphic encryption based algorithm finished clustering a large data set in approximately 66 seconds.

III. PROPOSED SYSTEM

We propose to classification the problem of unstructured data and providing the security to that data using the fuzzy c-means algorithm. We can use both the structured and unstructured data for the system and our system in compatible with the both of the data. Furthermore, we devise a privacy-preserving high-order fuzzy technique to enhance the efficiency by employing the cloud computing

A. Advantages of Proposed System

To cluster heterogeneous data in IoT, we present a high-order fuzzy c-means algorithm by extending the conventional fuzzy c-means algorithm to the high-order tensor space. In the high-order fuzzy c-means, we use the tensor distance to measure the distance between each pair of objects for avoiding the coincident clustering result.

To protect the private data when employing the cloud servers to improve the clustering efficiency for massive heterogeneous data in IoT, we propose a privacy-preserving high-order fuzzy c-means algorithm by using the BGV scheme to encrypt the original data

The BGV scheme does not support the division operations and exponential operations that are used in the membership matrix updating function of the high-order fuzzy c-means algorithm. To address this problem, we use the multi-Taylor technique to approximate the membership matrix updating function to a polynomial function.

IV. SYSTEM ARCHITECTURE

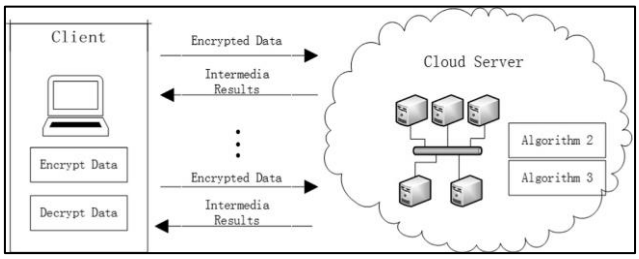

Fig. 1: System Architecture

V. CONCLUSION

We proposed a high-order fuzzy c-means algorithm to cluster heterogeneous data in IoT/cloud. Furthermore, we devise a privacy-preserving high-order fuzzy technique to enhance the efficiency by employing the cloud computing. BGV is used to protect the private data when performing the high-order fuzzy c-means on cloud. The idea of this paper is motivated by our previous work of PPHOPCM for multimedia data. However, there are at least two differences between PPHOPCM and PPHOFCM. First, PPHOFCM utilized tensor distance, instead of Euclide an metric, to capture the correlations of heterogeneous data. Second, PPHOFM utilizes the multi variable Taylor technique to transform the membership matrix updating function to a polynomial function to remove the division and exponentiation operations that are not supported by BGV while PPHOPCM uses the Taylor technique. Experiments implied that PPHOFCM out performs PPHOPCM for clustering heterogeneous data in IoT. Furthermore, our scheme is high scalable. In this paper, we focused on the heterogeneous data clustering for IoT. However, our schemes are not robust to noisy data which exists in IoT. Therefore, we will improve our schemes further to cluster heterogeneous data with noises more effectively.

REFERENCES

[1] European Network and Information Security Agency. Cloud computingsecurity risk assessment. https://www.enisa.europa.eu/activities/risk-management/files/deliverables/cloud-computing-risk assessment.

[2] Darcy A. Davis, Nitesh V. Chawla, Nicholas Blumm, Nicholas Chris-takis, and Albert-L ´ aszl ´ o Barabasi. Predicting individual disease risk based on medical history. In Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM ’08, pages 769–778, Napa Valley, California, USA, 2008.

[3] U.S. Dept. of Health & Human Services. Standards for privacy ofindividually identifiable health information, final rule, 45 cfr, pt 160–164. http://www.hhs.gov/sites/default/files/introduction.pdf, 2002.

[4] Jaideep Vaidya and Chris Clifton. Privacy-preserving k-means clusteringover vertically partitioned data. In Proceedings of the Ninth ACMSIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’03, pages 206–215, New York, NY, USA, 2003. ACM.

[5] Geetha Jagannathan and Rebecca N. Wright. Privacy-preserving dis-tributed k-means clustering over arbitrarily partitioned data. In Pro-ceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, KDD ’05, pages 593–599, New York, NY, USA, 2005. ACM.

[6] Paul Bunn and Rafail Ostrovsky. Secure two-party k-means clustering.In Proceedings of the 14th ACM Conference on Computer and Com-munications Security, CCS ’07, pages 486–497, New York, NY, USA,2007. ACM.

[7] Mahir Can Doganay, Thomas B. Pedersen, Y ¨ ucel Saygin, Erkay Savas¸,and Albert Levi. Distributed privacy preserving k-means clustering with additive secret sharing. In Proceedings of the 2008 International Workshop on Privacy and Anonymity in Information Society, PAIS '08, pages 3–11, New York, NY, USA, 2008. ACM.
[8] Jun Sakuma and Shigenobu Kobayashi. Large-scale k-means clusteringwith user-centric privacy-preservation. Knowledge and Information Systems, 25(2):253–279, 2009.
[9] Xun Yi and Yanchun Zhang. Equally contributory privacy-preserving k-means clustering over vertically partitioned data. Inf. Syst., 38(1):97– 107, March 2013.
[10] Rakesh Agrawal and Ramakrishnan Srikant. Privacy-preserving data mining. SIGMOD Rec., 29(2):439–450, May 2000.

