SeCCA: Homomorphic Encryption Based Privacy Preservation Scheme for Biclustering Algorithm

Shokofeh VahidianSadegh
Goethe University Frankfurt
Frankfurt, Germany

Lena Wiese
Goethe University Frankfurt
Frankfurt, Germany

Michael Brenner
Leibniz Universitaet Hannover
Hannover, Germany

34th Crypto Day, 9/10 June 2022

Massive amounts of newly generated gene expression data have been used to discover new insights that improve our understanding of the underlying biological processes (Vincent, Pierre & Pierre (2019)). In Yizong & M (2000), the concept of bicluster refers to a subset of genes and a subset of conditions with a high similarity score, which measures the coherence of the genes and conditions in the bicluster and also returns the list of biclusters for the given data set.

Moreover, as individual's biomedical data are sensitive, finding practical approaches to protect them is of great concern. Due to the increased threat of data breach, recently tremendous efforts have been made to encrypt private and personal genomic data files and also keep data encrypted during analysis, although there is a noticeable lack of studies in the area of processing data by biclustering algorithms that are privacy-preserved. Homomorphic encryption with certain operations (additions and/or multiplications) is able to handle sensitive genomic data by allowing data to remain encrypted even during computation.

In our work, we proved that homomorphic encryption operations can be applied directly on biclustering algorithms – particularly Cheng and Church algorithm. Secure Cheng and Church algorithm (SeCCA) is a privacy-preserving version of the original algorithm that implemented and tested with adjustable parameters on a real-world data set (yeast Saccharomyces cerevisiae cell cycle). As a proof of concept, we compare the result of biclusters from the Cheng and Church algorithm with SeCCA by external evaluation measure to clarify applicability of homomorphic encryption operations in biclustering algorithms. As the first study in this domain, our study demonstrates the feasibility of homomorphic encryption operations in gene expression analysis to achieve privacy-preserving biclustering algorithms. Although on the downside, we observe significantly increased execution time compared to the original algorithm. Besides, we experienced limitations of the homomorphic encryption implementation for the evaluation of conditional branching on encrypted data.

crypto day matters 34 (2022)
Beiträge vom 34. Kryptotag, 09.-10. Juni 2022, online.
doi:10.18420/cdm-2022-34-04

References

Branders Vincent, Schaus Pierre & Dupont Pierre (2019). Identifying gene-specific subgroups: an alternative to biclustering. *BMC bioinformatics* **20**(1), 1–13.

Cheng Yizong & Church George M (2000). Biclustering of expression data. In *Ismb*, volume 8, 93–103.