



International Quantitative Research and Applications Conference 2018
| 5 - 8 August 2018 | Kuching | Sarawak | MALAYSIA |

"Mathematical Challenges in Big Data Analytics"

Data Privacy: A New Approach and a Case Study in Cluster Analysis

Abdellah Salhi

Department of Mathematical Sciences,
University of Essex,
Wivenhoe Park, Colchester CO4 3SQ, UK
as@essex.ac.uk

Data Privacy (DP) has always been an issue in data analysis. This is more so today than ever before because of the advanced tools available to take advantage of data for all sorts of reasons including unethical. It has, therefore, become one of the big challenges that Big Data has thrown about in recent years. A number of attempts at dealing with DP and confidentiality preservation have been made. They mainly rely on data encoding, homomorphic encryption in particular, and other mathematical devices that allow datasets to be worked on in place of others for the benefit of getting the same or equivalent solutions. They do, however, have limitations often due to the high dimensionality of these datasets and their extremely large volumes. The curse of dimensionality and volume are of course inherent to the concept of Big Data.

In my talk, I will suggest a new approach that relies on complexity theory and NP-Completeness in particular. We describe this approach and illustrate it on a very common problem in data science, namely clustering. Results will be provided and discussed.

Relevant references

Taylor, C., and Salhi, A., "On Partitioning Multivariate Self-Affine Time Series", IEEE Transactions on Evolutionary Computation, 21(6), pp.845-862, (2017) DOI10.1109/TEVC.2017.2688521.

Selamoglu, B.I. and Salhi, A. "The Plant Propagation Algorithm for Discrete Optimisation: The Case of the Travelling Salesman Problem," in Nature-Inspired Computation in Engineering, Xin-She Yang (Editor), Studies in Computational Intelligence 637, pp. 43-61, 2016.

Salhi, A., and Toreyen, O., "A game theory-based multi-agent system for expensive optimisation problems". In Tenne,Y., Goh,C.-K.(Eds) Computational Intelligence in Optimization, pp.211-232, Springer-Verlag, 2010.

Salhi, A., Lausen, B., Rohmatul, M., and Toreyen, O., "A game theory framework for clustering", In Proceedings of the 6th IMT-GT Conference on Mathematics, Statistics and its Applications: ICMSA'2010, 2010.

Sokmen, O., "Cluster Analysis and its Application to the Question of Turkey's Accession to the European Union", Masters' Thesis, University of Essex, UK, 2011.

Wu, D., Haven, J., "Using homomorphic encryption for large scale statistical analysis", FHE-SI-Report, 2012, http://crypto.stanford.edu/~dwu4/papers/FHE-SI_Report.pdf

Craig, A., "A fully homomorphic encryption system". PhD thesis, Stanford University, 2009. crypto.stanford.edu/craig

Garey, M., and Johnson, D.S., "Computers and Intractability: A Guide to the Theory of NP-Completeness", Series of Books in Mathematical Sciences, W.H.Freeman and Co., NY, USA, 1979

Lenstra, J. K. and Rinnooy Kan, A. H. G., "Some Simple Applications of the Travelling Salesman Problem", Operational Research Quarterly, 26(4), pp.717-733, 1975.

Rivest, R.L., Adleman, L., and Dertouzos, M.L., "On data banks and privacy homomorphisms", Journal of Foundation of Secure Computation, pp.169-179, 1978.