Achieving Both Valid and Secure Logistic Regression Analysis on Aggregated Data from Different Private Sources

Yuval Nardi, Technion-Israel Institute of Technology, Haifa, Israel
ynardi@ie.technion.ac.il
Stephen Fienberg, Carnegie Mellon University, Pittsburgh, PA USA
fienberg@stat.cmu.edu
Robert J. Hall, Carnegie Mellon University, Pittsburgh, PA USA
rjhall@cs.cmu.edu

August, 2012
LARC-TR-04-12

LARC Technical Report Series: http://smu.edu.sg/centres/larc/larc-technical-reports-series/

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Carnegie Mellon University

ABSTRACT

Preserving the privacy of individual databases when carrying out statistical calculations has a relatively long history in statistics and had been the focus of much recent attention in machine learning. In this paper, we present a protocol for fitting a logistic regression when the data are held by separate parties - without actually combining information sources - by exploiting results from the literature on multi-party secure computation. Our protocol provides only the final result of the calculation compared with other methods that share intermediate values and thus present an opportunity for compromise of values in the individual databases. Our paper has two themes: (1) the development of a secure protocol for computing the logistic parameters, and a demonstration of its performances in practice, and (2) the presentation of an amended protocol that speeds up the computation of the logistic function. We illustrate the nature of the calculations and their accuracy using an extract of data from the Current Population Survey divided between two parties. Throughout, we build our protocol from existing cryptographic primitives, thus the novelty is in designing a concrete procedure for private computation of the logistic regression MLE rather than to propose new cryptographic constructions.