WebGLORE: a Web service for Grid LOgistic REgression

Wenchao Jiang\textsuperscript{1}, Pinghao Li\textsuperscript{1}, Shuang Wang\textsuperscript{2}, Yuan Wu\textsuperscript{3}, Meng Xue\textsuperscript{1}, Lucila Ohno-Machado\textsuperscript{2}, Xiaqian Jiang\textsuperscript{2}

\textsuperscript{1}Shanghai Jiaotong University, Shanghai, China
\textsuperscript{2}Division of Biomedical Informatics, University of California, San Diego, La Jolla, CA
\textsuperscript{3}Department of Biostatistics & Bioinformatics, Duke University, Durham, NC

Supported by AHRQ grant R01HS019913, an EDM Forum grant, and NIH Grant U54HL108460 to the University of California San Diego

Introduction

Gathering a large number of observations is important to ensure sufficient statistical power for a variety of studies, including randomized clinical trials. The wide adoption of electronic health records has made it technically possible to pull information from geographically distributed institutes efficiently. However, it is not easy to enable cross-institution collaboration involving patient-level data sharing because of the increasing concern that inappropriate leakage of personal information might put the privacy of individuals at risk. We previously addressed this problem by building a web service that implements a distributed logistic regression (LR) model using horizontally-partitioned data (i.e., using the same type of data that are available at different sites but not transmitting these data to a central site), but did not have a web service for this purpose.

Methods

We have previously studied distributed LR models and provided a theoretical justification for their soundness. We developed a software as a service (SAAS) package that allows data analysts to easily construct models across different sites using only their web browsers, without installing any software locally. The foundation of this work is based on Grid Logistic REgression (GLORE)\textsuperscript{[1]}, which uses aggregated partial results [e.g., covariance matrices that are transmitted to a central site] to synthetize a LR model. We developed a signed communication protocol between Servlets and Applets and state-of-the-art AJAX technology to support users’ experiences with dynamic webpage reflection. There is no installation requirement and the software can be easily deployed on any operation system and the most popular web browsers.

Results

We compared WebGLORE with a centralized LR model using a biomedical dataset. We split the data and simulated online collaboration. The results showed no differences in terms of Areas Under the ROC Curve (AUC), calibration, and attribute level statistics (e.g., odds ratios, z-statistics, etc.). In addition, our model had reduced computational cost because the computation was shared by participants. These results show the reliability and usability of WebGLORE as a promising tool for privacy-preserving exploratory analysis.

References
