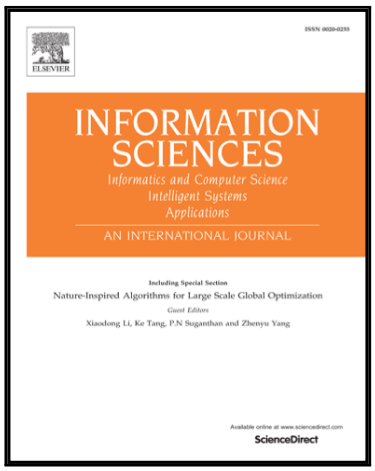

Jo-DPMF: Differentially Private Matrix Factorization Learning through Joint Optimization

Feng Zhang, Victor E. Lee, Kim-Kwang Raymond Choo

PII: S0020-0255(18)30595-4
DOI: 10.1016/j.ins.2018.07.070
Reference: INS 13837

To appear in: *Information Sciences*

Received date: 22 November 2017
Revised date: 24 July 2018
Accepted date: 26 July 2018

Please cite this article as: Feng Zhang, Victor E. Lee, Kim-Kwang Raymond Choo, Jo-DPMF: Differentially Private Matrix Factorization Learning through Joint Optimization, *Information Sciences* (2018), doi: 10.1016/j.ins.2018.07.070

Jo-DPMF: Differentially Private Matrix Factorization Learning through Joint Optimization

Feng Zhang[a,b,*], Victor E. Lee[c], Kim-Kwang Raymond Choo[d,e,a]

[a] *School of Computer Science, China University of Geosciences, Wuhan 430074, China*
[b] *Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430074, China*
[c] *GraphSQL Inc., Mountain View, CA 94043 , USA*
[d] *Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, TX 78249, USA*
[e] *Information Assurance Research Group, University of South Australia, Adelaide, SA 5095, Australia*

Abstract

Stochastic gradient descent (SGD) is a widely-used technique to implement matrix factorization. SGD-based matrix factorization involves many iterative computations. Therefore, according to the sequential composition theory of differential privacy, conventional implementation strategies of differentially private matrix factorization may lead to significant error accumulation, no matter whether the Laplace noise is added to the original matrix or to the factorized matrices. In fact, the implementation of differentially private matrix factorization is so challenging that results proposed to date have the problem of inefficient privacy and data utility. In this paper, we employ the objective perturbation method to address the challenge; this method dramatically alleviates error accumulation by perturbing the objective function instead of perturbing the results. Our method outperforms the state-of-the-art methods since it only requires a scalar noise rather than a vector noise to achieve the same magnitude of privacy. Furthermore, our method may learn the resulted matrices by joint optimization, which follows the conventional learning procedure of SGD and optimizes its convergence speed and accuracy as much as possible. In addition to the differential privacy guarantee, we also empirically show the way that the novel model works together with k-coRating, a k-anonymity-like privacy preserving model, to enhance data utility.

* Corresponding author
Email address: jeff.f.zhang@gmail.com (Feng Zhang)