Muhammad Kamran, Ehsan Ullah Munir, 2018. On the role of optimization algorithms in ownership-preserving data mining. *Frontiers of Information Technology & Electronic Engineering*, 19(2):151-164. https://doi.org/10.1631/FITEE.1601479

On the role of optimization algorithms in ownership-preserving data mining

Key words: Information security; Optimization; Digital rights; Watermarking

Corresponding author: Muhammad KAMRAN
E-mail: muhammad.kamran@ciitwah.edu.pk
ORCID: https://orcid.org/0000-0002- 6639-5688

Motivation

1. Watermarking of outsourced statistical datasets poses a unique challenge, as insertion of watermark into a feature of the dataset may change the predictive ability of that feature.

2. Consequently, the knowledge extracted from the dataset may be invalid.

3. Therefore, watermarking must be optimized while recognizing some usability constraints.

Main idea

1. The quality of the embedded watermark should ensure the following two major requirements:

(1) it must be robust against all possible malicious attacks for the deterioration of the watermark;

(2) a watermark must not damage the original dataset and the data usability must be ensured after watermark insertion.

Method

1. Watermarking is modeled as a constraint optimization problem and GA, GP, PSO, and MINLP approaches have been used to optimize watermark encoding.
2. Watermarking is optimized such that the predictive ability (classification potential) of a feature does not change after encoding.
3. The proposed methodologies are tested on a large number of datasets with varying usability constraints.
4. The robustness of the optimized watermark against various malicious attacks is validated.

Major results

1. The embedded watermark was found to be robust against various malicious attacks like tuple insertion, tuple deletion, and tuple updation attacks.

2. The embedded watermark also ensured the data quality that resulted in preserving of classification accuracy and the output of various feature selection schemes.

Conclusions

1. In this paper, the appropriateness of using optimization schemes for ownership-preserving data mining has been examined.

2. For this purpose, the insertion of a watermark has been modeled as an optimization problem with these objectives:

   (1) preserving the classification potential of high-ranking features;

   (2) identifying the maximum available bandwidth while ensuring the usability constraints;

   (3) maximizing the watermark robustness by using the available bandwidth.