



INDERSCIENCE *Online*

The online platform for Inderscience Publishers journal content

[Home](#) [Browse](#)
[Inderscience Publishers](#)
[Orders](#)
[Authors](#)
[Librarians](#)

[Advanced search](#)

Home > International Journal of Intelligent Information and Database Systems > List of Issues > Volume 13, Issue 2-4 > An Earth mover's distance-based undersam ...

[< Previous article](#)
[Next article >](#)

An Earth mover's distance-based undersampling approach for handling class-imbalanced data

Gillala Rekha, V. Krishna Reddy, Amit Kumar Tyagi

<https://doi.org/10.1504/IJIIDS.2020.109463>

Published online 26 August 2020

[Abstract](#)
[PDF](#)

Abstract

Imbalanced datasets typically make prediction accuracy difficult. Most of the real-world data are imbalanced in nature. The traditional classifiers assume a well-balanced class distribution for training data but in practical datasets show up an imbalance, thus obscure a classifier and degrade its capability to learn from such imbalanced datasets. Data pre-processing approaches address this concern by using either random undersampling or oversampling techniques. In this paper, we introduce Earth mover's distance (EMD), as a similarity measure, to find the samples similar in nature and eliminate them as redundant from the dataset. Earth mover's distance has received a lot of attention in wide areas such as computer vision, image retrieval, machine learning, etc. The Earth mover's distance-based undersampling approach provides a solution at the data level to eliminate the redundant instances in majority samples without any loss of valuable information. This method is implemented with five conventional classifiers and one ensemble technique respectively, like C4.5 decision tree (DT), k-nearest neighbour (k-NN), multilayer perceptron (MLP), support vector machine (SVM), naive Bayes (NB) and AdaBoost technique. The proposed method yields a superior performance on 21 datasets from Keel repository.

Keywords: class imbalance, classification, data pre-processing, sampling technique, Earth mover's distance, EMD

[Purchase this article](#)
[Subscribe this journal](#)

Click 'Add to cart' to add this article to the shopping cart. This article price is \$40.00. You may review the list of added articles prior to making the actual purchase on the shopping cart page.

International Journal of Intelligent Information and Database Systems



Print ISSN: 1751-5858 **Online ISSN:** 1751-5866

- [Current issue](#)
- [List of issues](#)
- [Subscribe](#)
- [Get TOC alerts](#)
- [About this journal](#)

Article / Chapter Tools

[Add to Favourites](#) |
 [Email to a Friend](#) |
 [Send to Citation Mgr](#) |
 [Track Citations](#)

Related Content Search

By Keyword

- class imbalance
- classification
- data pre-processing
- sampling technique
- Earth mover's distance
- EMD

By Author

- Gillala Rekha
- V. Krishna Reddy
- Amit Kumar Tyagi

[Most Read](#)
[Most Cited](#)

K-means** – a fast and efficient K-means algorithms

Applying distance sorting selection in differential evolution

Moral hazard resolved by common-knowledge in principal-agent model

Personality modelling and sentiment analysis on Chinese micro-blog posts

Solve the IRP problem with an improved discrete differential evolution algorithm

[See More](#)

Keep in touch:     



[Inderscience Online](#)
[Advanced Search](#)
[Browse](#)



[Inderscience Publishers](#)
[Subscribe](#)
[Authors](#)
[Librarians](#)



[Inderscience Submissions](#)
[Submissions Guidelines](#)
[Submit an Article](#)