Deep Regression Modeling for Imbalanced and Incomplete Time-Series Data

Publisher: IEEE Cite This PDF

Murtadha D. Hssayeni ; Behnaz Ghoraani **All Authors**

Abstract

Abstract:
During the collection of time-series data, many reasons lead to imbalanced and incomplete datasets. Consequently, it becomes challenging to develop deep convolutional mod... **View more**

Metadata

Abstract:
During the collection of time-series data, many reasons lead to imbalanced and incomplete datasets. Consequently, it becomes challenging to develop deep convolutional models without suffering from overfitting. Our objective in this paper was to investigate an emerging but rather underutilized framework of Conditional Generative Adversarial Networks (cGANs) for improving deep regression models for time-series data with an imbalanced and incomplete distribution. First, we investigated the potential of using a vanilla cGAN as a data imputation to improve the generalizability of the developed models to unseen data in such datasets. Next, we proposed a modified cGAN architecture with improved extrapolation and generalizability of the regression models. Our investigations used an imbalanced synthetic non-stationary dataset, a real-world dataset in Parkinson's disease (PD) application domain, and one publicly-available dataset for Negative Affect (NA) estimation. We found that vanilla cGAN failed to generate realistic time-series data due to severe mode collapse, limiting its application as a data imputation for imbalanced and incomplete data. Importantly, the proposed cGAN framework significantly improved extrapolation and generalizability for the prediction of regression scores with an average improvement of 56%, 34%, and 18%, respectively, in mean absolute error for the synthetic, PD, and NA datasets when compared with traditional Convolutional Neural Networks. The codes are publicly available on Github.

Published in: IEEE Transactions on Emerging Topics in Computational Intelligence (Volume: 8 , Issue: 6, December 2024)

Page(s): 3767 - 3778

DOI: 10.1109/TETCI.2024.3372435

Date of Publication: 19 March 2024

Publisher: IEEE

Electronic ISSN: 2471-285X

Funding Agency:

Contents

I. Introduction

Imbalanced datasets commonly exist in real-world applications. This paper addresses challenges with such imbalanced datasets, specifically in the biomedical domain when estimating disease severity scores (i.e., regression scores). The data is usually collected from a small group of patients, resulting in a data representation with a geometric distribution [1], [2], [3], [4]. This distribution means long-tailed normal distribution with some scores or classes representing the distribution head and other rare ones representing the tail. Hence, the data collection leads to imbalanced and, in most cases, incomplete datasets as the recruited patients may not represent the entire range of the disease severity score. Deep-learning models suffer overfitting when trained on these constrained datasets that violate the parsimony principle [5]. Overfitted models have poor generalizability to testing samples with underrepresented or unseen regression scores [6]. Multiple approaches are developed to handle overfitting in classification problems with cross-sectional data such as images [7]. One approach is based on Generative Adversarial Networks (GANs) that have emerged in numerous applications [8], [9]. For instance, GANs have been used to learn and generate new samples as a data augmentation method. The generated samples were then used in the training process in addition to the real samples. Unlike cross-sectional data, the overfitting challenges of deep regression problems in imbalanced time-series signal data have been overlooked. This is while a practical solution to this problem is essential in health monitoring applications such as over-time monitoring of human health or disease severity [10].

Authors

Figures

References

Keywords

Metrics

Footnotes

More Like This

SIGAN-CNN: Convolutional Neural Network Based Stepwise Improving Generative Adversarial Network for Time Series Classification of Small Sample Size
IEEE Access
Published: 2024

3D Convolutional Neural Network for Hyperspectral Image Classification Using Generative Adversarial Network
2020 13th International Conference on Intelligent Computation Technology and Automation (ICICTA)
Published: 2020

Show More