Studies in Computational Intelligence

Volume 989

Series Editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland

The series "Studies in Computational Intelligence" (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

Indexed by SCOPUS, DBLP, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

Rahul Kumar Sevakula · Nishchal K. Verma

Improving Classifier Generalization

Real-Time Machine Learning based Applications

Springer

Rahul Kumar Sevakula
Signal Processing
Whoop Inc
Boston, MA, USA

Nishchal K. Verma
Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, India

ISSN 1860-949X ISSN 1860-9503 (electronic)
Studies in Computational Intelligence
ISBN 978-981-19-5072-8 ISBN 978-981-19-5073-5 (eBook)
https://doi.org/10.1007/978-981-19-5073-5

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Classification algorithms form the basis of decision-making in most pattern recognition problems, e.g. image recognition, speech and speaker recognition, iris recognition, and spam mail detection. With the horizon of their applications expanding at a fast pace, the need for further research has only increased. This fact becomes particularly true because (a) each application poses its own set of challenges and (b) one would always find a classifier with a particular improvisation that best suits the situation. No matter which classification approach is used, generalization is an important aspect. Generalization essentially indicates how well the trained classifier works in real time, i.e. on unseen test data.

This monograph begins with the fundamentals of classifiers, bias-variance tradeoff, statistical learning theory (SLT), probably approximate correct (PAC) framework, maximum margin classifiers, and popular methods which improve generalization like regularization, boosting, transfer learning, dropout in deep learning, etc. Furthermore, the monograph solves four independent problems that have great relevance for certain real-time applications.

The first part of the monograph aims at finding classifiers which exhibit extremely low variance. Classification algorithms are traditionally designed to simultaneously reduce errors caused by bias as well as variance. In many situations, low variance becomes extremely crucial for getting tangible classification solutions and even slight overfitting can have serious consequences on the test results. Classifiers with low variance have two main advantages: (1) the classifier statistically manages to keep the test errors close to the training error, and (2) the classifier learns effectively even with a small number of samples. The monograph introduces a class of classifiers called Majority Vote Point Classifier (MVPC), which on account of the lower Vapnik Chervonenkis (VC) dimension can exhibit lower variance than even linear classifiers. The monograph contributes by estimating a trend for the MVP classifier's VC dimension, and validates its low variance on two real-time problems.

The monograph then focuses on the real-time application of condition-based monitoring of machines using acoustic and vibration measurements. Signal data acquired from machines are often found to change with time, wear and tear, and subsequent repair of the machine. Classifiers are typically trained to perform the decision-making procedure during fault diagnosis/detection. Since data may change with time, low generalization error is essential to avoid overfitting during classification. Therefore, MVPC is seen to be best suited for this situation. However, MVPC has a limitation that it may not be able to fit the data sufficiently, and may have high training errors. The monograph presents a novel framework for pattern recognition, where novel procedures for optimal data source (sensitive position) identification, data acquisition, and feature selection are tailored to give the best possible training performance with the MVP classifier. The understanding here is that if MVP gives low training error, real-time fault diagnosis of machines becomes feasible with consistent accuracy. The introduced framework was experimentally implemented and tested for an air compressor condition monitoring application; associated real-time experiments showed a significant improvement in the reliability of fault detection.

The third part of the monograph focuses on dealing with class noise in the Fuzzy Support Vector Machine (FSVM) classifier. FSVM is considered to be a significant addition over soft margin SVM like C-SVM, for the former can guard against outlier sensitivity and the latter cannot. The ability of FSVM to absorb outliers strongly depends on how well the training samples are assigned fuzzy membership values (MVs). Traditionally, the membership functions (MFs) used for FSVM were custom-made for applications, and MFs used for one could in general not be used for others. To overcome the limitation, General Purpose Membership Functions (GPMFs) are defined in this monograph as those MFs which can universally be used for multiple applications, and which allow FSVM to statistically perform better than C-SVM. The monograph contributes to the GPMF literature in two stages. Firstly with help of convex hulls, it presents a few limitations that FSVM faces while treating all samples of a class with a single MF. Further, it recommends differential treatment of data by categorizing them into two fuzzy sets: one containing possible non-outliers and the other containing possible outliers. While possible outliers are modeled with a normal MF, possible non-outliers are recommended to have a constant MV of ‘1’. The chapter then introduces novel GPMFs which use clustering-based techniques to detect possible outliers, and use Hausdorff Distance and pt-set to characterize those possible outliers. To establish conclusions, the introduced GPMFs are thoroughly evaluated and statistically compared with earlier GPMFs on numerous real-world benchmark datasets. The results show that proposed GPMFs not only perform significantly better in treating class noise, but also execute with efficient run time complexity.

Finally, a novel scheme to introduce deep learning in Fuzzy Rule-based classifiers (FRCs) is presented. FRCs have gained prominence for their unique ability of giving good classification performance, and allowing existing expert knowledge to be used conjointly with training data. Recent innovations in Deep Neural Networks are allowing researchers to tackle some very complex problems with improved theoretical and empirical justifications, e.g. image classification and audio classification. The monograph presents a scheme to incorporate stacked denoising sparse autoencoders within the FRC framework. While stacking of denoising sparse autoencoders helps learn the complex non-linear relationships among data and represent the input data in a reduced compact feature space, the framework built toward FRC allows users to input expert knowledge to the system. To make denoising sparse autoencoders learn more effectively, data pre-processing strategies have been proposed. Further, to improve the classification performance and rule reduction performance of the FRC, three fine-tuning strategies have also been proposed. The proposed framework is tested across real-world benchmark datasets, and an elaborate comparison across literature shows that proposed methods are capable of building FRCs that provide state-of-the-art accuracies and/or a few rules, as per the user's demand.

The monograph ends with an epilogue, on the use of autoencoders in transfer learning, tumor classification, and condition monitoring problems. Furthermore, pertaining to the research contributions made herein, the directions for possible future work have also been discussed.

Boston, USA Rahul Kumar Sevakula
Kanpur, India Nishchal K. Verma

Contents

About the Authors

Dr. Rahul Kumar Sevakula has over 10 years of research experience in machine learning (ML) and deep learning (DL). He received his Bachelor's degree from the National Institute of Technology (NIT) Warangal, India in 2009 and later his Ph.D. degree from the Indian Institute of Technology (IIT) Kanpur, India in 2017. He is currently a Sr. Research Scientist at Whoop, and his research interests lie at the intersection of ML, physiological signals, cardiovascular health monitoring (medicine) and wearables. Prior to joining Whoop, he was an Instructor (junior research faculty) at Harvard Medical School and Massachusetts General Hospital, USA, and a Data Scientist at IBM India. He has filed multiple patent disclosures and has published over 45 research papers in international peer-reviewed journals and conferences. He is also a reviewer for several journals of national and international repute.

Dr. Nishchal K. Verma is a Professor in the Department of Electrical Engineering at Indian Institute of Technology (IIT) Kanpur, India. Dr. Verma's research interest falls in Artificial Intelligence (AI) related theories and its practical applications to inter-disciplinary domains like machine learning, deep learning, computer vision, prognosis and health management, bioinformatics, cyber-physical systems, complex and highly non-linear systems modeling, clustering, and classifications, etc. He has published more than 250 research papers in peer-reviewed reputed conferences and journals along with 4 books (edited/co-authored) in the field of AI. He has 20+ years of experience in the field of AI. He is currently serving as Associate Editor/ Editorial Board Member of various reputed journals and conferences. He has also developed several AI-related key technologies for The BOEING Company, USA.

Acronyms

AE	Autoencoders
ANN	Artificial Neural Network
CBM	Condition-based Maintenance
CIL	Class Imbalance Learning
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DBSCAN	Density-based Spatial Clustering of Applications with Noise
DEC	Differential Error Cost
DFT	Discrete Fourier Transform
DL	Deep Learning
DNN	Deep Neural Network
DR	Dataset Rejection
DRL	Dataset Retrieval
DSAE	Denoising Sparse Autoencoders
EM	Expectation Maximization
FCM	Fuzzy C-Means
FDA	Fisher’s Discriminant Analysis
FFN	Feed Forward Network
FR	Feature Ranking
FRC	Fuzzy Rule Classifier
FS	Feature Selection
FSVM	Fuzzy Support Vector Machine
GA	Genetic Algorithm
GM	Gaussian Mixture
GMM	Gaussian Mixture Model
GPMF	General Purpose Membership Function
HD	Hausdorff distance
ITER	Individual Training Error Reduction
LIV	Leakage Inlet Valve
LOV	Leakage Outlet Valve
MDL	Minimum Descriptive Length

MF	Membership Function
MV	Membership Value
MVP	Majority Vote Point
MWT	Morlet Wavelet Transform
NDM	Normalized distance of means
OPTICS	Ordering Points To Identify the Clustering Structure
PRI	Parameter Range Identification
RBF	Radial basis function
RBM	Restricted Boltzman Machine
RM	Ratio of Means
RNN	Recurrent Neural Network
SAE	Stacked Autoencoders
SC	Separation Count
SDM	Standard deviation of means
SDSAE	Stacked Denoising Sparse Autoencoders
SGD	Stochastic Gradient Descent
SLT	Statistical Learning Theory
SPA	Sensitive Position Analysis
SSAE	Stacked Sparse Autoencoders
SVM	Support Vector Machine
VC	Vapnik Chervonenkis
WPT	Wavelet Packet Transform
ZCR	Zero Crossing Rate

Symbols

$\mathbb{R}$	Space of real numbers
Ω	Sample space or Metric space
S	A subset of Ω
H	Hypothesis space
Π_H	Growth function of H
υ, Υ	VC dimension
x	Scalar variable used for defining functions
$\mathbf{x}$	Input vector
$\mathbf{X}$	Input data matrix containing m samples and n features
$\mathbf{x}_i$	Feature vector of ith sample
$\mathbf{x}_j$	Vector containing jth feature values of all samples
x_{ij}	jth feature value of ith sample in $\mathbf{X}$
$\mathbf{y}$	Class label vector corresponding to $\mathbf{X}$
y_i	Class label of ith sample
$\mathbf{O}$	Desired output matrix in neural network
$\mathbf{o}_i$	Desired output vector of ith sample
o_{ij}	jth feature value of ith sample in $\mathbf{O}$
i	Data sample number
j	Feature number
m	Number of data samples
n	Number of features
p	pth class in classification problem
P	Number of classes in classification problem
$\mathbf{w}$	Weight vector
$\mathbf{W}$	Weight matrix
b	Bias variable
α, β, λ	Classifier and statistical test related parameters
ζ_i	ith Lagrange multiplier
C	Cost value in C-SVM
γ	RBF kernel parameter
ξ_i	Slack variable of ith sample in C-SVM

s_i	Membership value of ith sample for FSVM
Δ	Margin between two classes w.r.t. a hyper-plane
δ	A small value above 0
r	Radius of sphere and ellipsoid
d	Distance metric
ε	Radius of ε neighborhood
τ	Threshold value
g	Fuzzy rule number i.e. gth fuzzy rule
G	Number of fuzzy rules
k	kth cluster or fuzzy region
K	Number of clusters and number of fuzzy regions
l	lth neuron layer in ANN
L	Number of layers in ANN
ρ	Sparsity parameter for sparsity constraint
N	Number of datasets
σ	Standard deviation
ν	Mean
θ	Parameters of GMM
μ	Membership function
E	Number of epochs
f	Generic function
R	Risk function
TSDSAE	Function representing the training of SDSAE

List of Figures

List of Tables