

Conferences > 2016 International Conferen...

A cluster-filter feature selection approach

Publisher: [IEEE](#) [Cite This](#) [PDF](#)

Vimal Kumar Dubey, Amit Kumar Saxena, Madan Madhaw Shrivastava [All Authors](#)

[2](#) Paper
[150](#) Citations
[Full Text Views](#)



Abstract

Document Sections

I. Introduction

Abstract:

In this paper, a feature selection method is presented for the multiclass data sets. This method is the hybridization of k-means clustering using cosine similarity as a distance measure and information Gain. In the method unsupervised Cosine Similarity is used for grouping of features i.e. K-means clustering is used to make a cluster of features and then

More Like This

[Accuracy Enhancement of Correlated Naive Bayes Method by Using Correlation Feature Selection \(CFS\) for Health Data Classification](#)
 2020 3rd International Conference on Information and Communications Technology (ICICT) | Published: 2020

[Random Forest Framework Customized to Handle Highly Correlated Variables: An](#)

Abstract

Document Sections

I. Introduction

II. Preliminaries

III. Proposed Method

IV. Datasets and Experiments

V. Results and Discussions

[Show Full Outline](#)

[Authors](#)

[Figures](#)

[References](#)

[Citations](#)

[Keywords](#)

[Metrics](#)

Abstract:

In this paper, a feature selection method is presented for the multiclass data sets. This method is the hybridization of k-means clustering using cosine similarity as a distance measure and information Gain. In the method unsupervised Cosine Similarity is used for grouping of features i.e. K-means clustering is used to make a cluster of features and then information gain is employed to select a most relevant feature from each cluster. The dataset with the selected feature is tested for classification accuracy with cross-validation approach. Three classifiers namely Naive Bayes (NB), K-Nearest Neighbor and Classification and Regression trees (CART) has been used as the base classifiers for getting classification accuracy. Obtained results are compared with filter-based feature selection technique (Information Gain).

Published in: 2016 International Conference on ICT in Business Industry & Government (ICTBIG)

Date of Conference: 18-19 November 2016

INSPEC Accession Number: 18792102

Date Added to IEEE Xplore: 06 April 2017

DOI: 10.1109/ICTBIG.2016.7892637

ISBN Information:

Publisher: IEEE

Conference Location: Indore, India

I. Introduction

Multiclass High-dimensional datasets classification is very problematic due to the property of the datasets that numbers of patterns are very less compared to a number of features. Curse of Dimensionality [7] plays a big role in the classification problem. Since classification is an indispensable part of data mining [1], machine learning [2] or pattern recognition [3], there must be some device for which classification should be good as much as possible with

[technology \(ICICAT\)](#)
 Published: 2020

[Random Forest Framework Customized to Handle Highly Correlated Variables: An Extensive Experimental Study Applied to Feature Selection in Genetic Data](#)
 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA) | Published: 2018

[Show More](#)