Chapter 22
A Combined Feature Selection Technique for Improving Classification Accuracy

S. Meganathan

iD https://orcid.org/0000-0003-4570-8259
SASTRA University (Deemed), India

A. Sumathi
SASTRA University (Deemed), India

Ahamed Lebbe Hanees
South Eastern University of Sri Lanka, Sri Lanka

ABSTRACT

Feature selection has become revenue to many research regions that manage machine learning and data mining since it allows the classifiers to be cost-efficient, time-saving, and more precise. In this chapter, the feature selection strategy is consolidating by utilizing the combined feature selection technique, specifically recursive feature elimination, chi-square, info-gain, and principal component analysis. Machine learning algorithms like logistic regression, random support vector machine, and decision trees are applied in three different datasets that are pre-processed with combined feature selection technique. Then these algorithms are ensembled using voting classifier. The improvement in accuracy of the classifiers is observed by the impact of the combined feature selection.

INTRODUCTION

Dimensionality reduction is the way toward eliminating repetitive or insignificant highlights from the first informational index. So the execution season of the classifier that measures the information decreases, likewise precision increments on the grounds that unessential highlights can incorporate noisy

data influencing the order exactness adversely. With include determination the umderstandability can be improved and cost of information dealing with decreases.

Feature selection, as a dimensionality decrease strategy, expects to pick a little subset of the significant highlights from the original features by eliminating unimportant, excess or noisy highlights. Feature selection, normally can prompt better learning execution, i.e., higher learning precision, lesser computational expense, and better model interpretability.

This work utilizes combined feature selection technique and Machine learning strategies for forecast with higher exactness rate. The proposed strategy has used with the machine learning method. The component determination system shows the significance of choice procedure. It gives the better precision for calculations with the diminished dataset.

We utilize three particular dataset namely, skin ailment dataset (1st dataset). The UCI facility provided the dataset. The dataset includes 22 Histopathological features and 12 clinical attributes. The target variable contains the characteristics from 1 to 6. 2nd dataset-> was acquired from UCI which is a diabetic dataset called Pima Diabetes dataset .3rd dataset->Diabetes dataset was acquired from kaggle. It contains 50 segments and 1 lakh records. Diabetes dataset maintains portrayal in an exceedingly twofold association that involves 0 and 1 which represents diabetes non-affected and diabetes affected independently.

The CFS is employed to select the rule credits. Four component assurance methods namely Principal component analysis, Recursive feature elimination, Chi-square and Info-gain are joined for performing the endeavour. This system performs diversely on the datasets. Joining the methodologies which are performing particularly would pass highest outcomes. So, selecting the customary features which are looked over all the four component assurance strategy records. Diabetes dataset maintains request during a twofold arrangement that contains 0 and 1 which represents diabetes non-affected and diabetes affected independently.

LITERATURE SURVEY

Xie and Wang (2011) introduced a new strategy called half breed highlight determination, it improves F-score value along with Sequential Forward Search (IFSFS). The research got the main F-score by separating courses of real numbers assessment and isolation among extra genuine number arrangement. The better F-score and Sequential Forward Search are merged as ideal segment during time spent component assurance, be that as it may, that improved F-score as channel methodology evaluation model, and, SFS as appraisal system covering procedure.

Aruna et al (2012) came out with a hybrid feature decision procedure explicitly IGSBFS (Information Gain and Sequential Backward Floating Search), which solidifies potential gains channels just as covers pick ideal part beginning with first rundown of abilities reliant upon a characteristic model of NB. Information Gain variable evaluator are used in the opposite assurance progressively or otherwise called as straight forward decision along with skimming forward decision (IGSBFS) (FS2). In IGSBFS, IG functions as channels to dispose of abundance features. The course of action precision of the offered, method put forward is 98.9% with 10 features.

Xie et al (2013) developed an alternate hybrid FS computation across F-score where critical attributes as per a disorder dataset. Incorporate subset age pertinent features, glancing through techniques, for instance, SBFS, ESFS, and SFFS and summarized F-score to survey meaning of every segment. These

Related Content

Genetic Adaptation of Level Sets Parameters for Medical Imaging Segmentation
Dário A.B. Oliveira, Raul Q. Feitosaand Mauro M. Correia (2010). *Biomedical Image Analysis and Machine Learning Technologies: Applications and Techniques (pp. 150-166).*
www.irma-international.org/chapter/genetic-adaptation-level-sets-parameters/39559

Image Focus Measure Based on Energy of High Frequency Components in S-Transform
Muhammad Tariq Mahmoodand Tae-Sun Choi (2013). *Image Processing: Concepts, Methodologies, Tools, and Applications (pp. 162-180).*
www.irma-international.org/chapter/image-focus-measure-based-energy/77544

Human Face Recognition using Gabor Based Kernel Entropy Component Analysis
Arindam Kar, Debotosh Bhattacharjee, Dipak Kumar Basu, Mita Nasipuriand Mahantapas Kundu (2012). *International Journal of Computer Vision and Image Processing (pp. 1-20).*
www.irma-international.org/article/human-face-recognition-using-gabor/74797

Multimedia Image Retrieval System by Combining CNN With Handcraft Features in Three Different Similarity Measures
Maher Alrahhaland Supreethi K.P. (2020). *International Journal of Computer Vision and Image Processing (pp. 1-23).*
www.irma-international.org/article/multimedia-image-retrieval-system-by-combining-cnn-with-handcraft-features-in-three-different-similarity-measures/245667

Faster Ranking Using Extrapolation Techniques
Muhammad Ali Norozi (2013). *Intelligent Computer Vision and Image Processing: Innovation, Application, and Design (pp. 179-198).*
www.irma-international.org/chapter/faster-ranking-using-extrapolation-techniques/77040