Доклади на Българската академия на науките
Comptes rendus de l'Académie bulgare des Sciences
Tome 75, No 7, 2022

ENGINEERING SCIENCES
Signal processing

ENHANCED FEATURE SELECTION METHOD USING WRAPPER-BASED RANDOM SEARCH STRATEGY AND MUTUAL INFORMATION FOR REMOTE SENSING IMAGE CLASSIFICATION

T. Gladima Nisia, S. Rajesh*

Received on December 14, 2021
Presented by Ch. Roumenin, Member of BAS, on March 29, 2022

Abstract

Remote sensing image classification is one of the useful image processing tasks and finds application in many real-life scenarios. Image feature is important and unavoidable in classification. It is impossible to get a clear classification result without proper features. So after feature extraction, selecting an efficient and relevant feature is inevitable. Thus, our system proposes a new way of selecting features by a wrapper-based method that works using a randomized search strategy. The process is done in an orderly manner. In each step, the features that contribute less to classification are rejected to bring out the most relevant features. Three steps are followed: (1) Randomized selection (2) Warm start and (3) Cool down. The Randomized selection selects relevant features based on the random search method from the full feature set. Among the selected features, the important features are selected, based on the Mutual Information using the Warm start. Some important features, missed out in Randomized selection, are picked up in Cool down. The system finds out 25% of the most relevant features with greater classification accuracy when compared with classification accuracy obtained using 100% features. The proposed system has been checked for its efficiency with the help of remote sensing-based datasets from the UCI repository and it is found to be more efficient than the other existing methods. The results produced with the selected features are of high accuracy and low computational cost.

Key words: remote sensing image classification, image processing, feature selection

DOI:10.7546/CRABS.2022.07.12

1. Introduction. The selection of features is the method of picking the most relevant features. Due to the importance of the selection process, it is implemented in many different areas, such as bioinformatics [1], text mining [2], processing of images [3], ecological modelling [4], and industrial fault diagnosis [5]. Li et al. [6] and Jovic et al. [7] presented a complete summary of feature selection applications. The method is broadly classified into three main categories: filter, wrapper, and embedded method. The Filter method selects features based on some criteria which do not use any machine learning techniques. Even though the filter-based method seems to be very easy and fast to implement, it produces low classification due to the lack of learners. The wrapper method uses learner independent methods for better accuracy that results in high computational cost. For example, graph-based approaches [8,9] are proposed as filter methods. Even though the features are selected they are not present with good quality. In the wrapper method, optimization algorithms are used to find the ideal feature set. But it consumes a lot of time for producing efficient features since the method uses the learner as a black box and assesses the learner's performance. The algorithm proposed in our method falls under the wrapper method category. The last category of feature selection method is the embedded method. This method works in a hybrid way by selecting the important features while training the learner. Decision tree and Least Absolute Shrinkage and Selection Operator (LASSO) models are typical examples. But this hybrid technique can be embedded in a certain number of machine learning models only. The implementation is hard for most of the models such as KNN [10].

The traditional wrapper method depends on conventional searching algorithms like branch & bound and sequential search [11,12]. But there are chances for features to get trapped in local optima. To avoid this drawback, random search-based optimization approaches are introduced [13,14]. These methods locate improved feature set with high computational costs. As a consequence, lightweight randomized selection algorithms become the recent trend. Randomized feature elimination algorithms and selection algorithms [15–17] are developed to find useful feature subsets. In this system, a novel method is proposed for an efficient random search-based wrapper method for feature selection by including three novel techniques to accelerate and enhance the feature selection process. The rest of the work is structured as follows: Section 2 delivers a methodology of the proposed feature selection method. In Section 3, the results are discussed. The last Section 4 concludes the paper.

2. Methodology. After the features are extracted, the important method of feature selection is undergone to obtain the most relevant features. They are important for the process of classification. The proposed method performs this selection in three steps: (1) Randomized selection, (2) Warm Start, and (3) Cool Down.

2.1. Randomized selection. The basic idea of randomized selection is very simple. The system computes the feature group size (represented as 'g') and selects 'g' number of features randomly for feature selection from the set, F to G. The learner then evaluates the accuracy of the features in set, G. Based on the random selection method, 'n' ($n < g$) number of features are added to X. Accuracy of the features in G and X is calculated. The feature set with high accuracy is selected for the next iteration. If the accuracy is not improved, then the old features will remain in the set G. The process is continued until the specified number of iterations.

Algorithm: randomized selection (F)
//Input: Features which are extracted, F
//Output: Selected feature, G
Compute feature group size, g;
Add random 'g' number of features to set, G;
while(sizeof (G) < g)
    Evaluate the accuracy, acc of G using learner;
    X = G; Then, remove 'n' features from X and Add 'n' features from F to X;
    Evaluate the accuracy, new_acc of X;
    if acc < new_acc
        G = X;
    else
        G remains the same;
    end-if
end-while
Return final selected features, G

2.2. Warm start. Since the features are selected based on the random search process in the previous step, the features now undergo another round of selection based on the feature importance. The proposed system makes sure that the most relevant features are included in the final set using some standard procedures. The process evaluates the feature set, G using the Mutual Information (MI). The features with larger MI values are considered the best features. The MI of two random variables is calculated by entropy and it measures the uncertainty of random variables and scales the total information shared by them effectively [18]. The MI is calculated using Eq. (1)

$$\mathrm{I}(\mathrm{X};\mathrm{Y}) = \sum_{\mathrm{y}\in\mathrm{Y}}\sum_{\mathrm{x}\in\mathrm{X}} \mathrm{P}(\mathrm{x},\mathrm{y})\log_2\frac{\mathrm{p}(\mathrm{x},\mathrm{y})}{\mathrm{p}(\mathrm{x})\mathrm{p}(\mathrm{y})} \tag{1}$$

where P(x, y) is the joint probability distribution function of X and Y. Let X be a discrete random variable and Y be the Class label. The MI value is calculated for all the features in set G and the 'm' number of features with larger MI values is placed in another set, E.

2.3. Cool down. Now, the features in set E remain as the final features. Since our first step itself is based on random search methodology, there is some probability of missing a few relevant features in the selection process from the set, F. So, in the cool down process the system performs the final check for useful features in F. If some features are found to be relevant, they are added to the final feature set, E_select.

Algorithm: cool_down (E, F)
//Input: Selected feature set, E and full feature set, F
//Output: Final feature set, E_select
Calculate MI for all the features in the set, F – E.
Fea_order = Arrange the features in descending order based on MI values.
E1 = Select 'p' number of features from Fea_order
X = E
i = 1
while (i<p)
    Evaluate the accuracy, acc of X using learner.
    E_select = Add i^{th} features from E1 to X;
    Evaluate the accuracy, new_acc of E_select;
    if acc < new_acc
        E_select = X;
    else
        E_select remains the same;
    end-if
    i = i +1;
end-while
Return final selected features, E_select

The E_select features are selected as final features of our proposed method and these features are sent for classification.

3. Results and discussions. The section discusses the validation and evaluation of the proposed method. All the experiments are implemented using Matlab 2020a. The system specifications are listed in Table 1.

Table 1

System specifications

Name	Detailed Settings
CPU	Intel(R) Core(TM) i5 1.80 GHz
RAM	8.0 GB
OS	Windows 10
Language	MATLAB 2020a

3.1. Dataset description. All the experiments are conducted on the UCI benchmark datasets. The details about the datasets are mentioned in Table 2.

Table 2

Benchmark datasets

S. No	Dataset	Number of Instances	Number of Features
1	Urban Land Cover	168	148
2	Forest types	326	27
3	Wilt	4889	6
4	Sat Images	6435	36
5	Abalone	4177	8
6	Breast cancer	286	9
7	Ionosphere	351	34
8	Spectrometer	531	102
9	Hill-Valley	606	101
10	Wine	178	13

3.2. Experimental results. The proposed system performs three steps for feature selection systematically. The first step implemented is Randomized selection. Initially, the feature group size, g is computed using Eq. (2). Our proposed system selects the 'g' number of features randomly for feature selection from the set, F. The learner implemented here is a gradient boosting tree. The algorithm is run with different iterations and the best number of iterations are found out for all the datasets. It is observed that the classification accuracy of most of the dataset remains constant after 100 iterations. So, each run of the algorithm implemented here consists of 100 iterations. The result of the Urban Land Cover dataset is shown for reference in Fig. 1. The selected features are

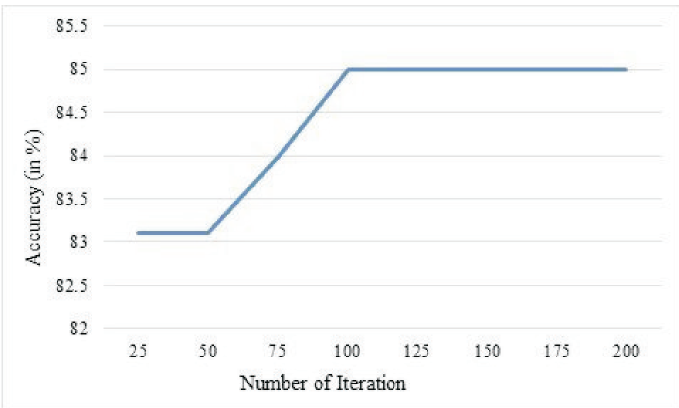

Fig. 1. Comparison against Number of Iterations and Classification Accuracy of Urban Land Cover dataset

stored in set G using a randomized selection algorithm. This system results in selecting 25% of the best features.

(2)
$$g = \left\lfloor \frac{\text{Total number of Features}}{4} \right\rfloor .$$

The second step of the proposed system is Warm start. Here, the MI value is calculated for all the features in set G using Eq. (1), and the 'm' number of features with larger MI values is placed in another set, E. At this point, the system results in selecting 12% of the best features.

(3)
$$m = \left\lfloor \frac{g}{2} \right\rfloor .$$

The third step implemented is cool down. The cool down process is implemented to find out the remaining useful features in the feature set, F. The 'm' value is assigned to 'p' also. Among 'p' features some features are found to be relevant and they are added to the final feature set, E_select. Finally, the system results in selecting 12% to 25% of best features from the feature set, F.

The results of the dataset are tabulated in Table 3. The system is observed to produce efficient classification accuracy. Also from the table, it is observed that the total number of features chosen is found to be within 12% to 25% of the total number of features.

Table 3

Results of our proposed method

S. No	Dataset	Classification Accuracy (in %)	Total Number of Features	Selected Number of Features
1	Urban Land Cover	83.5	148	27
2	Forest types	85	27	5
3	Wilt	81.2	6	3
4	Sat Images	79.9	36	8
5	Abalone	86.4	8	4
6	Breast cancer	82.7	9	4
7	Ionosphere	84.1	34	8
8	Spectrometer	80.3	102	19
9	Hill-Valley	82.4	101	18
10	Wine	85.5	13	5

4. Conclusion. The system implements the wrapper-based random search method. The wrapper method uses learners for selecting features that are followed by random based methodology. The randomized selection step results in selecting 25% of the best features. The warm start step results in selecting 12% of the best features among the 25% features. The cool start step adds up to another 12% of the features to the final set. Finally, the method results in selecting 12% to 25% of the best features from the whole set of features. From the implementation of our proposed method, it is clearly understood that the method can improve the classification results using the selected features.

REFERENCES

[1] Wang L. P., Y. L. Wang, Q. Chang (2016) Feature selection methods for big data bioinformatics: A survey from the search perspective, Methods, **111**, 21–31.

[2] Mladenić D. (2011) Feature selection in text mining. In: Encyclopedia of Machine Learning, (Eds C. Sammut and G. J. Webb), Boston, MA, USA: Springer, pp. 406–410.

[3] Huang K., S. Aviyente (2008) Wavelet feature selection for image classification, IEEE Trans. Image Process., **17**(9), 1709–1720.

[4] Tracy J. L., A. Trabucco, A. M. Lawing, J. T. Giermakowski, M. Tchakerian et al. (2018) Random subset feature selection for ecological niche models of wildfire activity in Western North America, Ecol. Modell., **383**, 52–68.

[5] Liu C., D. X. Jiang, W. G. Yang (2014) Global geometric similarity scheme for feature selection in fault diagnosis, Exp. Syst. Applicat., **41**(8), 3585–3595.

[6] Li J. D., K. W. Cheng, S. H. Wang, F. Morstatter, R. P. Trevino et al. (2017) Feature selection: A data perspective, ACM Comput. Sur., **50**(6), 94.

[7] Jovic A., K. Brkic, N. Bogunovic (2015) A review of feature selection methods with applications. In: Proc. 2015 38th Int. Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 1200–1205.

[8] Zhang Z. H., E. R. Hancock (2011) A graph-based approach to feature selection. In: Proc. Int. Workshop on Graph-Based Representations in Pattern Recognition, Münster, Berlin, Germany, 205–214.

[9] Das A. K., S. Goswami, A. Chakrabarti, B. Chakraborty (2017) A new hybrid feature selection approach using feature association map for supervised and unsupervised classification, Exp. Syst. Appl., **88**, 81–94.

[10] Fonti V., E. Belitser (2017) Feature Selection Using LASSO, `https://pdfs.semanticscholar.org/24ac/d159910658223209433cf4cbe3414264de39.pdf`.

[11] Nakariyakul S., D. P. Casasent (2009) An improvement on floating search algorithms for feature subset selection, Patt. Recognit., **42**(9), 1932–1940.

[12] Russell S., P. Norvig (2009) Artificial Intelligence: A Modern Approach, New York, NY, USA: Prentice Hall.

[13] Xue B., M. J. Zhang, W. N. Browne, X. Yao (2016) A survey on evolutionary computation approaches to feature selection, IEEE Trans. Evol. Comput., **20**(4), 606–626.

[14] Ghamisi P., J. A. Benediktsson (2015) Feature selection based on hybridization of genetic algorithm and particle swarm optimization, IEEE Geosci. Remote Sens. Lett., **12**(2), 309–313.

[15] Stracuzzi D. J., P. E. Utgoff (2004) Randomized variable elimination, J. Mach. Learn. Res., **5**, 1331–1362.

[16] Saha S., S. Rajasekaran, R. Ramprasad (2015) Novel randomized feature selection algorithms, Int. J. Found. Comput. Sci., **26**(3), 321–341.

[17] Brankovic A., A. Falsone, M. Prandini, L. Piroddi (2018) A feature selection and classification algorithm based on randomized extraction of model populations, IEEE Trans. Cybernet., **48**(4), 1151–1162.

[18] Su X., F. Liu (2018) A Survey for Study of Feature Selection Based on Mutual Information. In: 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) 2018 Sep 23, IEEE Amsterdam, Netherlands, DOI: 10.1109/WHISPERS.2018.8746913.

Department of Computer Science
and Engineering
AAA College of Engineering and Technology
Sivakasi – 626123, Tamil Nadu, India
e-mail: `gladimab@gmail.com`

**Department of Computer Science*
and Engineering
Mepco Schlenk Engineering College
Sivakasi – 626005, Tamil Nadu, India
e-mail: `srajesh@mepcoeng.ac.in`