International Journal of Trend in Scientific Research and Development, Volume 1(2), ISSN: 2456-6470
www.ijtsrd.com

Hybrid Optimization Techniques to Improve Feature Selection in Image Classification Techniques

Uma G. Kartha
Lecturer in Electronics, Government Polytechnic College, Kothamangalam, Kerala, India

ABSTRACT
In data mining, feature selection is an important phase in the data preprocessing process. Finding a subset of features such that a classification model constructed just using this subset has superior prediction accuracy than a model constructed solely using the full set of features is the problem of feature selection. Two hybrid strategies for feature selection are presented here. In any case, the best features are chosen by combining current feature selection methods or developing new ones from scratch. Classification models based on five classifiers are then built using the reduced dataset. An area under the receiver operating characteristic (ROC) curve (AUC) performance metric was used to measure classification accuracy in this study. It has been demonstrated empirically that the proposed methods can increase the performance of existing feature selection methods.

Keywords*: Hybid, Features selecton, Image, Wrapper, PSO,GWO*

Introduction
Pre-processing in machine learning and data mining relies heavily on dimensionality reduction and anomaly detection. Real-world datasets, on the other hand, tend to be influenced by both errors and redundant variables. Also, data selection has two basic issues: removing aberrant patterns or reducing the number of dimensions in the data. When the original variables must be preserved, feature selection is a typical dimensionality reduction strategy.

In pattern recognition, machine learning, and computer vision, feature selection is critical. the decrease of dimensionality and the improvement of machine learning performance and process comprehension are the major goals of feature selection. In order to locate the optimal subsets, the exhaustive search method is the only method that can ensure this, however its computational time complexity is exponential.

Feature selection approaches can be broken down into three groups based on the method of selection: These include techniques such as filtering, wrapping, and embedding. Filter approaches do not necessitate additional learning algorithms, whereas wrapper techniques must. Compared to filtering, the wrapper approach has higher processing costs, as well as an over-fitting risk, which makes it less efficient. But in the case of embedded approaches, the method of selecting features is integrated into the training process of the model(s), and then an ideal set of features is generated by optimising the objective function.

Feature selection
PSO and GWO are used to pick features in the hybrid technique, which benefits from good convergence and faster search time.

- Particle swarm optimization

An initial random population, known as particles, is used as a starting point for the PSO's optimization process. Each particle has n-dimensional solutions to a single problem and a single procedure.

- Grey wolf optimization

GWO [23] is an optimization method that mimics the hunting behaviour of a grey wolf pack. Processes like prey, track, encircle, and attack make up GWO. This is the four-tiered structure of the grey wolf's hierarchy: alpha α, beta β, omega ω, and delta δ. When it comes to attacks, the alpha α is in charge of making decisions, while beta β provides support. The beta is responsible for carrying out the decisions of the alphas at the lower levels of the hierarchy and reporting back to the alphas. Other wolves, such as sentinels and scouts, fill in for the omega, who serves as the scapegoat. alpha and beta, aided by delta knowledge, now control omega.

Four methods of feature selection are outlined in the literature, including filter, wrapper, embedding, and hybrid. The next section provides a brief summary of

the FS approaches. As shown in Fig. 1, there are many various ways to select features.

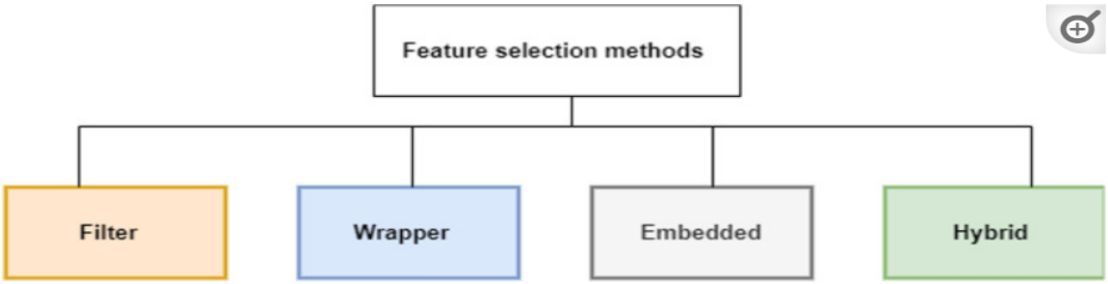

Fig. 1 shows a separation of feature selection techniques.

Review of Literature

Instead of learning features as the model is built, wrapper techniques employ induction algorithms to pick suitable feature subsets based on the classifier's accuracy (Kohavi and John 1997). Filtering metrics can be employed as a first step, followed by a wrapper or embedded technique, in a hybrid approach.

Computational and statistical scalability are strong strengths for filtering algorithms (Inza et al. 2007). They don't matter what kind of algorithm is utilised to teach them. Filtering results can be used in a variety of mining methods. The strategies of filtering, however, are reliant on the unique statistical measurements they use and so biassed towards the metric that they utilise. Using them in the learning model does not improve classification accuracy since they do not produce improved feature subsets. Furthermore, selecting the correct filtering metric is a separate problem.

Multivariate filtering is a subset of univariate filtering. Additionally, filtering methods can also be categorised into feature weighting and feature subset selection approaches, respectively. Relevance information is considered in univariate approaches, but feature dependencies are ignored. As a result, multivariate methods are slower and less scalable than univariate ones (Canedo et al. 2013)

Ranking aggregation methods are employed to increase diversity and take advantage of the advantages of individual approaches. They also avoid exaggerating the value of their products. Rank aggregation, on the other hand, is a separate issue because different filtering metrics may rank the same features differently (Waad et al. 2014).

Objectives

- To investigate a hybrid approach to feature selection
- To learn about the concept of feature selection.
- In order to study the many classifications of features,

Research Methodology

A research technique is a standardised approach to gathering data, analysing that data, and drawing conclusions from the findings of a study. A strategy for conducting a study is known as a research technique. Research can be described as the systematic collection and analysis of facts and information for the progress of knowledge in any field. The study's objective is to develop intellectual and practical solutions to difficulties via the application of systematic methodologies. Based on secondary sources such as books, journals, academic articles, government publications and printed and online reference resources, this study is descriptive rather than prescriptive in character.

Result and Discussion

In terms of computational convenience, filter techniques are more convenient than wrappers since they are not tied to the learning machine and are also faster than wrappers.

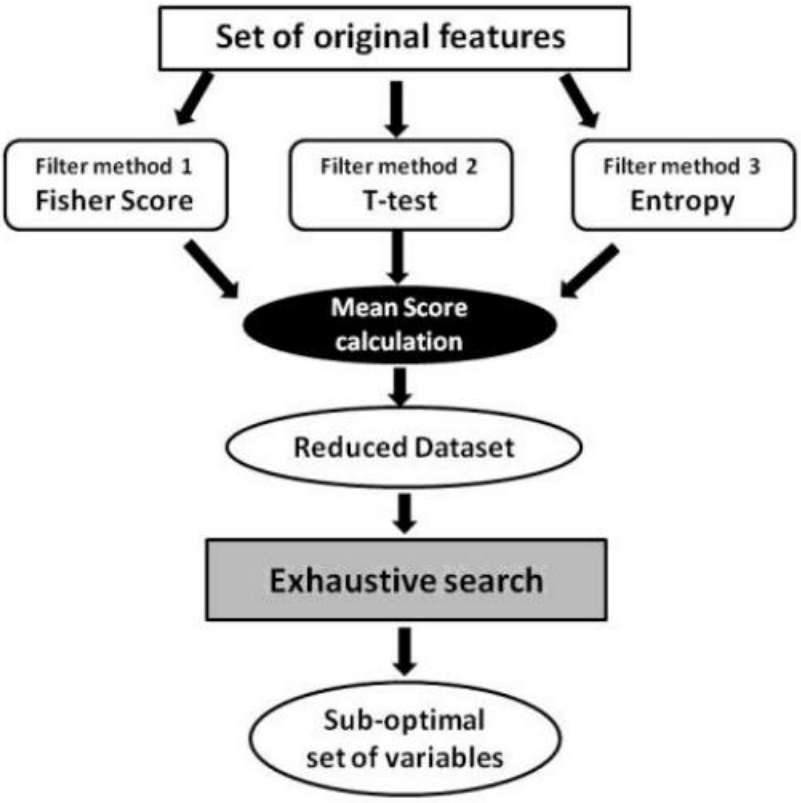

Figure 2 shows a generic diagram outlining the procedure under consideration.

The suggested method is a hybrid filter-wrapper method because exhaustive search picks the variables that most affect the target on the basis of the classifier performance.

To select the most relevant characteristics from a huge data set is a major issue in feature selection. Because certain features are so closely tied to the topic at hand while others are not, it might be difficult to find particular links and come to a decision when dealing with a huge dataset. The selection outcome would be affected if all of the features were picked.

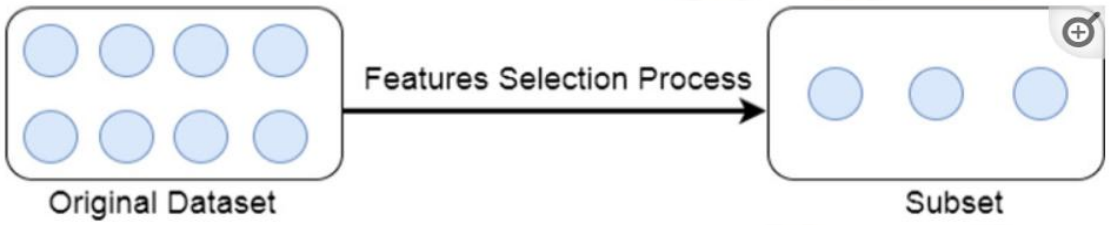

Figure 3: Concept of feature selection

As a result, in order to come up with the most effective solution, only include characteristics that are directly relevant to the situation at hand. If any of the elements of the analysis can affect the conclusion, which will lead to inaccurate results or take a long time, they should be avoided. Fig. 3 depicts the idea of reducing the number of attributes in a big dataset during feature selection.

The term hybrid refers to the idea of merging weaker methods to create a more effective and efficient model. Both feature selection and model training can be carried out simultaneously using the hybrid approaches.

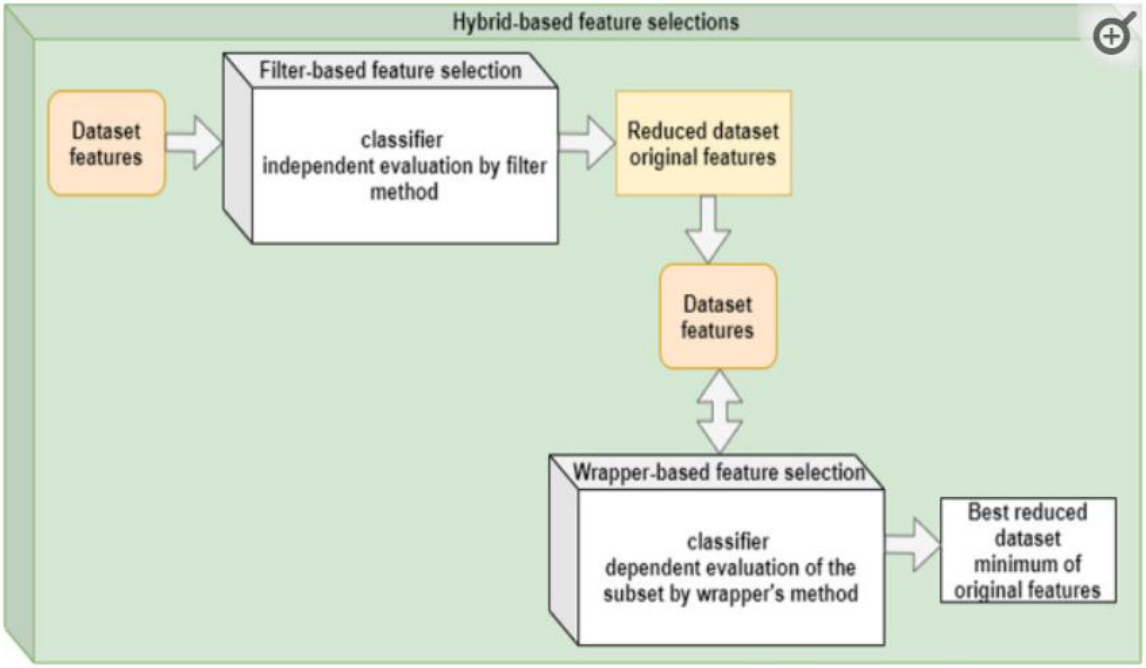

Figure 4 shows a feature selection hybrid technique that combines the filter and wrapper methods.

There are numerous benefits to employing hybrid approaches, including excellent accuracy and performance as well as optimal computational complexity and flexible models. Filter and wrapper selection methods can be used together in hybrid methods as shown in Fig. 4.

Conclusion

A filter-wrapper approach for feature selection is shown. The primary idea is to combine different filter methods for feature selection and then use an exhaustive search to find the optimum combination of remaining characteristics in an acceptable amount of time. The relevant features are selected using the hybrid PSO/GWO feature selection approach, and the Parallel MSVM method applies them to the features that have been selected. The hybrid feature selection technique has the advantages of high search efficiency and good convergence. PSO and GWO feature selection methods can't compete with the hybrid feature selection method. This reveals that the suggested hybrid feature selection using the Parallel MSVM approach has 99.15% accuracy, whereas the existing RF method has 94.1% accuracy, based on this data. In the future, deep learning will be used to increase the method's performance.

References

[1] Kohavi R, John GH (1997) Wrappers for feature subset selection. Artif Intell 97(1–2):273–324

[2] Inza I, Larranaga P, Saeys Y (2007) A review of feature selection techniques in bioinformatics. Bioinformatics 23(19):2507–2517

[3] Canedo VB, Marono NS, Betanzos AA (2013) A review of feature selection methods on synthetic data. Knowl Inform Syst 34:483–519

[4] Waad B, Ghazi BM, Mohamed L, LARODEC I, LARIME E (2014) A new feature selection technique applied to credit scoring data using a rank aggregation approach based on: optimization, genetic algorithm and similarity. In: Knowledge discovery process and methods to enhance organisational performance, pp 347–376

[5] Cateni, S., Colla, V., Vannucci, M. A fuzzy system for combining different outliers detection methods. Proceedings of IASTED International Conference on Artificial Intelligence and Applications, AIA 2009, pp. 87-93.

[6] Cateni, S., Colla, V., Vannucci, M. A fuzzy logic based method for outliers detection. Proceedings of IASTED International Conference on Artificial Intelligence and Applications, AIA 2007, pp. 561-566.

[7] Cateni, S., Colla, V., Vannucci, M., Vannocci, M. A procedure for building reduced reliable training datasets from real-world data. Proceedings of the IASTED International

Conference on Artificial Intelligence and Applications, AIA 2014, pp. 393-399.

[8] Wang S. and Zhu J. Variable selection for model-based high dimensional clustering and its application on microarray data.Biometrics, 64. pp.440-448, June 2008.

[9] Kwak, N. and Choi C.H. Input feature selection for classification problems. IEEE trans. on neural networks, Vol. 13, pp.143-159, 2002.

[10] V. Bolón-Canedo, N. Sánchez-Marono, A. Alonso-Betanzos, J. M. Benítez, and F. Herrera, "A review of microarray datasets and applied feature selection methods," Information Sciences, vol. 282, pp. 111–135, 2014.

[11] M. Alzaqebah and S. Abdullah, "An adaptive artificial bee colony and late-acceptance hill-climbing algorithm for examination timetabling," Journal of Scheduling, vol. 17, no. 3, pp. 249-262, 2014.

[12] M. Alzaqebah and S. Abdullah, "Hybrid bee colony optimization for examination timetabling problems," Computers & Operations Research, vol. 54, pp. 142–154, 2014.

[13] J. Friedman, T. Hastie, and R. Tibshirani, "The elements of statistical learning," Springer series in statistics New York, vol. 1, no. 10, 2001.

[14] C.-L. Huang et al. A GA-based feature selection and parameters optimization for support vector machines Expert Syst. Appl. (2006)

[15] D. Wang et al.t-test feature selection approach based on term frequency for text categorization Pattern Recognit. Lett.(2014)

[16] S. S. Keerthi and C.-J. Lin, ''Asymptotic behaviors of support vector machines with Gaussian kernel,'' Neural Comput., vol. 15, no. 7, pp. 1667–1689, Jul. 2003.

[17] K. Chin, S. DeVries, J. Fridlyand, P. T. Spellman, R. Roydasgupta,W. L. Kuo, A. Lapuk, R. M. Neve, Z. Qian, T. Ryder, and F.

[18] Chen, ''Genomic and transcriptional aberrations linked to breast cancer pathophysiologies,'' Cancer cell, vol. 10, no. 6, pp. 529–541, 2006.

[19] D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd,P. Tamayo, A. A. Renshaw, A. V. D'Amico, J. P. Richie, E. S. Lander,

[20] M. Loda, P. W. Kantoff, T. R. Golub, and W. R. Sellers, ''Gene expression correlates of clinical prostate cancer behavior,'' Cancer Cell, vol. 1, no. 2, pp. 203–209, Mar. 2002.