ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

HYBRID APPROACHES FEATURE SUBSET SELECTION: A REVIEW

Ankur Singh Bist*, Neha Pandey
* KIET Ghaziabad
TCS Noida

ABSTRACT

To overcome the weakness of filter and wrapper approaches, many researchers combined both the methods together. The hybrid approaches are intended to be computationally more effective than wrapper approach as well as yielding higher accuracy than filter approach.

KEYWORDS: Filter Approach & Wrapper Approach

INTRODUCTION

Hybrid Approach

Xing *et al.*, (2001) successfully applied feature selection methods (using a hybrid of filter and wrapper approaches) to a classification problem in molecular biology involving only 72 data points in a 7130 dimensional space. They also investigated regularization methods as an alternative to feature selection, and showed that feature selection methods were preferable in the problem they tackled.

Das, (2001) proposed a hybrid approach **boosting based hybrid for feature selection** (BBHFS) incorporates some of the features of wrapper methods into a fast filter method for feature selection. He combined the strengths of both filter and wrapper approaches while reducing their weaknesses. He incorporates the natural stopping criteria of wrapper approach into filter method to overcome the computational expense of wrapper method. The experimental results showed that BBHFS performs better than wrapper methods on the DNA dataset using Naive Bayes and on the Chess dataset using ID3.

Guyon and Elisseeff, (2003) gave an introduction to variable and feature selection and focussed mainly on constructing and selecting subsets of features that are useful to build a good predictor. They recommend using any linear predictor (e.g. a linear SVM) and select variables in two alternate ways: (1) with a variable ranking method using correlation coefficient or mutual information; (2) with a nested subset selection method performing forward or backward selection or with multiplicative updates. They also provides a better definition of the objective function, feature construction, feature ranking, multivariate feature selection, efficient search methods, and feature validity assessment methods.

Oh *et al* (2004) proposed a hybrid GA design to solve the feature selection problem with improved capability than the other conventional algorithms. They hybridized local search operations into the simple GA. Local search operations are devised and embedded in hybrid GAs to fine tune the search. The hybrid GAs showed better convergence properties compared to the classical GA. They demonstrated that the proposed hybrid approach outperforms other algorithms in terms of classification accuracy and the computation time particularly for large-sized problems. Another advantage offered by the hybridized approach was the acquisition of subset size control.

Sikora and Piramuthu, (2005) introduced a wrapper approach to feature selection based on Hausdroff distance measure and presented the framework for efficient genetic algorithm based feature selection techniques in data mining. The experimental results confirmed that prediction accuracy and computational efficiency improved highly after combining both the approaches (filter and wrapper).

Uncu and Turksen, (2007) proposed a novel feature selection approach that combine wrapper and filter approach in order to identify the significant input variables in systems with continuous domains. It utilizes concept of functional dependency, correlation coefficients and k-nearest neighbourhood (KNN) method.

Xiao et al., (2008) introduced a novel embedded feature selection method called ESFS which is inspired by the SFS approach. ESFS method not only selects the most relevant features but also perform classification without any need of an extra classifier. ESFS performance was better than the traditional filter method, like Fisher algorithm. In addition, ESFS provides better classification accuracy than representative state of the art classifiers, such as neural networks, or decision trees.

Foithong *et al.*, (2012) introduced a novel feature selection method based on the hybrid model. It uses mutual information criterion for the selection of the candidate feature set. Then wrapper approach searches into the space of candidate feature subsets to select a proper subset that suits the learning algorithm. Using filter approach with wrapper reduced the computational cost and eliminated the issue of local maxima of wrapper search. They compared their technique with other representative methods. According to the results, this approach outperformed the other methods not only in classification accuracy, but also with respect to the number of features selected.

FEATURE SELECTION USING GENETIC ALGORITHM

Genetic Algorithm is a stochastic global search method that mimics the metaphor of natural biological evolution. GA work with a set of candidate solutions called a population and the GA obtains the optimal solution after a series of iterative computations. There follows some salient methods of feature subset selection using GA:

Forrest, (1993) explained the genetic algorithm as a computational model of biological evolution. He emphasized that with an appropriate measure of fitness, GA has potential for solving problems, making models, optimizing a function or determining proper order of a sequence. He also analysed GA mathematically to explain how GA works and how best to use them.

Man *et al.*, (1996) in his paper 'Genetic Algorithms: Concepts and Applications' introduced GA as an optimization tool. He outlined the features of GA in terms of the genetic functionality of operators, the problem formulation, the inherent capability of GA for solving complex and conflicting problems, as well as its various practical applications.

Raymer *et al.*, (2000) presented a new approach where GA was used simultaneously for feature extraction, feature selection and classifier training. The genetic algorithm optimized a vector of feature weights used to scale the individual features and then this technique was combined with the k nearest neighbour classification rule. A comparison was also done with sequential floating forward feature selection, and linear discriminant analysis.

Whitley, (2001) presented an overview of evolutionary algorithms covering genetic algorithms, evolutionary strategies, genetic programming and evolutionary programming. Gray codes, bit representation and real-valued representations were also discussed for parameter optimization problems.

Salcedo-Sanz *et al.*, (2002) proposed a feature selection method based on genetic optimization. They developed a a novel Genetic Algorithm (GA) operator which fixes the number of selected features and hence reduces the size of the search space and improves the GA performance and convergence. this genetic operator is known as *m-features* operator.

Tan *et al.*, (2008) combined various existing feature selection method to take advantage of multiple feature selection criteria and find small subsets of features with better classification performance. There are a number of feature selection algorithms with different selection criteria. But no single criterion is best for all applications. These small feature subsets are then fed to GA in the second stage. Then the GA will try to search an optimal feature subset which performs better than each single individual feature selection algorithm does.

Kumar and Jyotishree, (2012) developed a Novel Encoding Scheme in Genetic Algorithms for Better Fitness. They studies different encoding techniques and their associated genetic operations and then proposed a new encoding scheme to overcome the limitations of existing encoding techniques.

EVALUATION USING SUPPORT VECTOR MACHINES

Support Vector Machines (SVM) was developed by **Vapnik (1979)** for image classification. Then **(Cortes & Vapnik, 1995)** introduced the soft margin hyperplane for non-separable data which made SVM

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

more applicable. SVM classifies data with different class labels by determining a set of support vectors that are members of the set of training inputs that outline a hyperplane in the feature space.

Multi-class SVM

Support Vector Machines is a well developed technique for binary classification. Extending it effectively for multi-class classification is still a research issue. Many works have dealt with this phenomenon for the last three decades; some of the recent works are as follows:

Scholkopf, (2000) briefly elaborated the basics of statistical learning theory, support vector machines, and kernel feature spaces and summarized some empirical findings and theoretical developments of support vector machines. They found that by using the kernels in SVM, the optimal margin classifier was turned into a classifier which became a serious competitor of high-performance classifiers.

Weston *et al.*, (2000) introduced a method of feature selection for Support Vector Machines. Gradient descent method was used to perform the search for key features and segregate dimension(s) with minimum weights. The removal of minimum weighted dimensions results in the performance improvement. The method is based upon finding features which minimize bounds on the leave-one-out error. This method outperforms the performance of various filter methods and speeds up SVMs for time critical applications.

Hsu and Lin, (2002) performed comparison of methods for multi-class support vector machines to effectively extend SVM from binary classification to multi-class classification. Basically there are two types of approaches for multi-class SVM. One is by constructing and combining various binary classifiers together, while in other all classes are considered at once. They gave decomposition implementations for two such all-together different methods and compare their performance with three methods based on binary classification: one-against-one, one-against-all and DAG. They experimentally show that one-against-one and DAG may be more suitable for practical use.

Mao, (2004) developed a discriminative function pruning analysis (DFPA) feature subset selection method in the context of support vector machines, to reduce computational cost and increase the speed during classification. This method was based on hybrid approach and combines the strength of both of the approaches i.e. filter and wrapper.

Mavroforakis and Theodoridis, (2006) demonstrated a Geometric Approach to Support Vector Machine (SVM) Classification. The geometric framework for the support vector machine (SVM) classification problem provides an intuitive ground for the understanding and the application of geometric optimization algorithms, leading to practical solutions of real world classification problems. Their result shows that existing geometric algorithms can be directly and practically applied to solve not only separable, but also nonseparable classification problems both accurately and efficiently.

SVM model combined with GA (GA-SVM)

GAs has been applied to find an optimal set of feature weights that improve classification accuracy and has proven to be an effective computational method for high dimensional data. SVMs are much more effective than other conventional nonparametric classifiers in terms of classification accuracy, computational time and stability to parameter setting. There are many works reported where GA is employed as an optimizer for the feature selection process of SVM classifier. Some salient works are as follows:

Fröhlich *et al.*, (2003) dealt with the problem of feature selection for SVMs by means of GAs. They used the theoretical bounds on the generalization error for SVMs to select the feature subset. There are a number of learning parameters that can be utilized in constructing SV machines for regression. The two most relevant are the insensitivity zone *e* and the penalty parameter *C*, which determines the trade-off between the training error and VC dimension of the model. GA was used to optimize the kernel parameters such as the regularization parameter 'C' of the SVM. They compared GA with Fisher Criterion Score, Relief-F and Recursive Feature Elimination methods using cross-validation. According to the results, GA is a recommendable alternative, if the number of features to select is not fixed.

Downs and Wang, (2004) employed GA to select data from the training set in such a way that a sequence of SVM solutions is obtained that moves towards the Bayes optimal solution. It improved the performance of SVM on a problem in object recognition. In their work they were dealing with a problem of robot that attempts to recognize objects from multiple sonar returns.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Li *et al.*, (2005) proposed a robust gene selection approach based on a hybrid between genetic algorithm and support vector machine. They hybridize SVM and GA due to the capability of SVM of handling high-dimensional dataset and the capability of GA to optimize the subset of key features. They analyzed the gene expression of malignant lymphoma, specifically the diffuse large B cell lymphoma (DLBCL). Their experimental results proved that GA coupled with SVM outperforms the marginal filters and a hybrid between genetic algorithm and K nearest neighbours by achieving higher accuracy (99%) for prediction of independent microarray samples.

Huang and Wang, (2006) stated that the kernel parameters setting for SVM in a training process impacts on the classification accuracy. The objective of their research is to simultaneously optimize the parameters and feature subset without degrading the classification accuracy. They presented a genetic algorithm approach for feature selection and parameters optimization to solve this kind of problem. They conducted experiments on several real-world datasets like iris, heart disease, breast cancer, vowel, sonar etc. using the proposed GA-based approach and the Grid algorithm, a traditional method of performing parameters searching. The proposed GA-based approach had good accuracy performance with less number of features as compared with the Grid algorithm.

Huerta *et al.*, (2006) introduced another GA-SVM approach for gene selection and classification of Microarray data. This approach was associated to a fuzzy logic based pre-filtering technique which allows reducing the data dimensionality largely by grouping similar genes. They used GA to identify potentially predictive gene subsets for which fitness is evaluated by a SVM classifier. This approach is a hybrid wrapper approach that combines GA with a SVM classifier. It is used both for selecting predictive genes and for final gene selection and classification. They applied this approach on the Leukemia and Colon cancer datasets and compared it with six previous methods. Their approach was able to obtain good classification accuracy.

Zhuo, (2008) introduced a Genetic Algorithm based wrapper feature selection method for classification of hyperspectral images using Support Vector Machine. In this work, he used GA to optimize both the feature subset, i.e. band subset, of hyper spectral data and SVM kernel parameters simultaneously. The experimental results show that GA-SVM method significantly increases the classification accuracy up to 92.51% and the number of bands also gets reduced and hence the computational cost reduced drastically.

Ding and Chen, (2010) addressed the problem of the classification of high dimensional data using intelligent optimization methods. They proposed two intelligent optimization methods, GA-FSSVM (Genetic Algorithm-Feature Selection Support Vector Machines) and PSO-FSSVM (Particle Swarm Optimization-Feature Selection Support Vector Machines) models where GA and PSO were two evolutionary computing approaches which were used to optimize the parameters of SVM. These approaches were also used to optimize the feature subset selection and thus increase the classification accuracy of SVM. According to their experimental results these methods provides better results than the traditional grid search approach and many other approaches.

Temitayo *et al.*, (2012) proposed a hybrid GA-SVM approach for efficient feature selection in e-mail classification. They developed a Genetic Algorithm-Support Vector Machine (GA-SVM) hybrid technique to optimize the feature subset selection and classification parameters for SVM classifier. In their work, they classified all the e-mails as spam(1) or legitimate(-1) and use two classifiers, SVM and GA-SVM to filter spams from the spam assassin dataset of e-mails. GA eliminates the redundant and irrelevant features in the dataset. Then SVM further optimize the feature subset. Their study shows that the hybrid GA-SVM achieves higher classification accuracy and lower computational time in comparison to SVM.

Wu *et al.*, (2012) proposed an approach which simultaneously combines feature selection and parameter setting, to improve the classification accuracy and decrease the computational time in classifying ultrasound breast tumour images. Texture and morphological features were used in this study to effectively distinguish between benign and malignant lesions of the breast. The feature selection step was implemented with the genetic algorithm which identifies the significant features used to evaluate the breast tumour images. GA was also used to find the near-optimal parameters C and γ of the SVM. According to the experimental result the proposed system can detect a malignant tumour with high probability.

Pal, (2013) made use of Digital Airborne Imaging Spectrometer hyperspectral data set to evaluate the

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

performance of a wrapper–filter genetic algorithm (GA) for feature selection. *k*-nearest neighbour (*k*-NN) and support vector machine (SVM) classifiers were used to obtain classification accuracy and used as a fitness function to GA. According to the experimental results computational cost is high using SVM in comparison to k-NN. Choice of filter algorithm seems to be having no significant effect on classification accuracy obtained using selected features.

CLASSIFICATION BY NEURAL NETWORK

Artificial neural networks (ANN's) have been used widely in many application areas in recent years. A neural network can be defined as a reasoning model based on the structure of human brain. Neural Network learns by adjusting the weights so as to be able to correctly classify the training data and hence, during testing phase, to classify unknown data. They needs long time for training but has a high tolerance to noisy and incomplete data. NN often produce better results when compared to other classifiers. Many works have applied neural networks for the last four decades; some important work includes:

Lippmann, (1987) gave an introduction to the field of artificial neural nets. He reviewed six important neural net models that are highly parallel building blocks. These nets illustrate neural net components that can be used to construct more complex systems for pattern classification. He also demonstrated that how simple neuron-like components can perform classification and clustering.

Sexton & Dorsey, (1998) compared backpropagation (BP) with the GA for Neural Network training, to overcome the limitations of gradient algorithms which are a variation of backpropogation. They found that a global search technique such as the GA reliably outperforms the commonly used BP algorithm as an alternative NN training technique.

Park *et al*., (2009) presented a new architecture and design methodology of a granular neural network. They developed a design strategy for radial basis function neural networks to reduce the dimensionality of input space over which receptive fields are formed. Context based clustering is used in the granulation of information for the development of RBFNNs. Genetic Algorithm is used for the evolutionary optimization of the input spaces. According to the experimental results fuzzification coefficient (m) plays an important role in the design of the model.

Ludwig and Nunes, (2010) proposed three maximum-margin training methods for Supervised Neural Networks. Two based on the back-propagation approach and a third one based on information theory for Multilayer Perceptron (MLP) binary classifiers. Both back-propagation methods are based on the Maximal Margin (MM) principle. The main idea is to compose a neural model by using neurons extracted from three other neural networks, each one previously trained by MICI, MMGDX, and Levenberg Marquard, respectively. The resulting neural network was named Assembled Neural Network (ASNN).

Kabir *et al.*, (2010) introduced a new feature selection method using neural network that is based on the wrapper approach. It automatically determined neural network architecture during the feature selection process. To reduce the redundancy in features and to build compact NN architectures a constructive approach for FS (CAFS) was developed. It uses correlation information to select less correlated features.

CONCLUSION

This paper gives a brief overview about Methods for HYBRID APPROACHES FEATURE SUBSET SELECTION. There are lots of advancements that are going on in this specific domain. Continuous evolution in this area has added various dimensions in base atoms of concerned area. This study will be helpful for those working in the area of Methods for HYBRID APPROACHES FEATURE SUBSET SELECTION.

REFERENCES

[1] Das, S. 2001. Filters, wrappers and a boosting-based hybrid for feature selection. In International Confernce in Machine Learning (ICML) Vol. 1, pp. 74-81.
[2] Dash, M. and Liu, H., 1997, Feature selection for classification. Intelligent Data Analysis: An Int'l J., vol. 1, Issue (3), pp. 131-156.
[3] Ding, S. and Chen, L. 2010. Intelligent Optimization Methods for High-Dimensional Data Classification for Support Vector Machines in Intelligent Information Management, Volume 2, Issue (6), pp. 354-364.
[4] Dogan, N. & Tanrikulu, Z. 2010. A comparative framework for evaluating classification algorithms. In Proceedings of the World Congress on Engineering Vol 1.
[5] Domeniconi, C.; Peng, J. & Gunopulos, D 2000. An adaptive metric for pattern

classification. Advances in Neural Information Processing Systems, Volume 13.
[6] Downs, T. & Wang, J. 2004. Improving support vector solutions by selecting a sequence of training subsets. In Intelligent Data Engineering and Automated Learning–IDEAL 2004 Volume 3177, pp. 696-701. Springer Berlin Heidelberg.
[7] Duan, K. B., & Keerthi, S. S. 2005. Which is the best multiclass SVM method? An empirical study. In Multiple Classifier Systems, Springer Berlin Heidelberg, Volume 3541, pp. 278-285.
[8] Duda, R. O., Hart, P. E., & Stork, D. G. 2012. Pattern classification. John Wiley & Sons.
[9] Dy, JG. and Brodley, C.E. 2000. FSS and order identification for unsupervised learning,In the Proceedings of the seventeenth International conference on Machine Learning, USA, pp. 247-254.
[10] Estévez, P. A.; Tesmer, M.; Perez, C. A. & Zurada, J. M. 2009. Normalized mutual information feature selection. Neural Networks, IEEE Transactions, Volume 20.
[11] Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Annals of eugenics, Volume 7, Issue (2), 179-188.
[12] Foithong, S.; Pinngern, O. & Attachoo, B. 2012. Feature subset selection wrapper based on mutual information and rough sets. Expert Systems with Applications, Volume 39 Issue (1), pp. 574-584.
[13] Forrest, S. 1993. Genetic algorithms: principles of natural selection applied to computation. Science, Volume 261 Issue (5123), pp. 872-878.
[14] Frohlich, H.; Chapelle, O. & Scholkopf, B. 2003. Feature selection for support vector machines by means of genetic algorithm. In Tools with Artificial Intelligence. Proceedings. 15th IEEE International Conference on IEEE, pp. 142-148.
[15] Guo, C. & Yang, X. 2011. A Programming of Genetic Algorithm in Matlab7. 0. Modern Applied Science, Volume 5, Issue (1), pp. 230.
[16] Gutlein, M.; Frank, E.; Hall, M. & Karwath, A. 2009. Large-scale attribute selection using wrappers. In Computational Intelligence and Data Mining. CIDM'09. IEEE Symposium on (pp. 332-339). IEEE.
[17] Guyon, I., & Elisseeff, A. 2003. An introduction to variable and feature selection. The Journal of Machine Learning Research, Volume 3, pp. 1157-1182.
[18] Hall, M. A. 1999. Correlation-based Feature Selection for Machine Learning Doctoral dissertation, The University of Waikato, Hamilton, NewZealand.
[19] Han, Y.; Park, K. & Lee, Y. K. 2011. Confident wrapper-type semi-supervised feature selection using an ensemble classifier. In Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 2011 2nd International Conferenceon (pp. 4581-4586). IEEE.
[20] Haupt, R. L. & Haupt, S. E. 2004. Practical genetic algorithms. John Wiley & Sons.
[21] Holmes, G.; Pfahringer, B.; Kirkby, R.; Frank, E. & Hall, M. 2002. Multiclass alternating decision trees. In Machine Learning: ECML 2002 (pp. 161-172). Springer Berlin Heidelberg.
[22] Huang, E.; Cheng, S. H.; Dressman, H.; Pittman, J.; Tsou, M. H.; Horng, C. F. & Huang, A. T. 2003. Gene expression predictors of breast cancer outcomes. The Lancet, Volume 361, Issue (9369), pp. 1590-1596.
[23] Hughes, G. F. 1968. On the mean accuracy of statistical pattern recognizers, in: IEEE Transactions Information Theory, IT, Volume 14, issue (1), pp. 55–63.
[24] John, G.H.; Kohavi, R. and Pfleger, K. 1994. Irrelevant features and the subset selection problem, In: Proceedings of the Eleventh International Conference on Machine Learning, Volume 94, pp. 121–129.
[25] Jung, I. & Wang, G. N. 2008. Pattern classification of back-propagation algorithm using exclusive connecting network. Int. Journal of Comp. Sci. Eng, Volume 2, Issue (2), pp. 76- 80.
[26] Kabir, M. & Islam, M. 2010. A new wrapper feature selection approach using neural network. Neurocomputing, Volume 73, Issue (16), pp. 3273-3283.
[27] Kohavi, R. and John, G. H. 1997, Wrappers for feature subset selection. Journal Artificial Intelligence – Special issue on relevance, Volume 97 Issue (1-2), pp 273- 324.
[28] Kotsiantis, S. B. 2007, Supervised Machine Learning: A Review of Classification Techniques, Informatica Volume 31, pp.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

249-268.

[29] Kotsiantis, S. B. & Pintelas, P. E. 2005. Logitboost of Simple Bayesian Classifier. Informatica (Slovenia), Volume 29, Issue (1), pp. 53-60.

[30] Kurian J. and Karunakaran, V. 2012, A Survey on Image Classification Methods International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), Volume 1, Issue (4).