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

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

FEATURE SUBSET SELECTION: A REVIEW

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

ABSTRACT

Image classification has earned enormous attention due to the advent of modern day applications involving image base information and now an extensive research has been carried out in this field. It is important to study earlier research and work done to know the basic knowledge and techniques used for classification of images. This paper comprises brief review of work done by researchers and scientist in the domain of feature subset selection.

KEYWORDS: Filter Approach & Wrapper Approach

INTRODUCTION

IMAGE CLASSIFICATION

Image classification refers to the labelling of images into one of a number of predefined categories. Classification includes image acquisition, image pre-processing, image feature extraction, image feature subset selection and classification using a classifier. A number of classification techniques have been developed over these years for image classification. Two main image classification methods are supervised and unsupervised classification. Some of the effective work done on image classification techniques is as follows:

Pal *et al*., (1998) proposed a method of generating class boundaries in n-dimensional data space for classifying patterns using Genetic Algorithm. He found it essential for feature space to be bound and discretized upto sufficiently small interval for being classified by GA. The effectiveness of the proposed method is extensively demonstrated on two sets of artificial data and real life data of iris, speech and satellite imagery. The result of the approach was compared with Bayes classifier, K-NN and MLP. He found GA to outperform the other classifier and yield performance comparable to the Bayes classifier. The paper further discussed the issues of generalization capability of the GA as well as the issues of removing redundancy.

Majumdar and Jayas, (2000) developed classification models by combining more than one feature sets (morphological, color, textural) to classify individual kernels of grains like Canada Western Red Spring (CWRS) wheat, Canada Western Amber Durum (CWAD) wheat, barley, oats, and rye. The mean accuracies (the average of the classification accuracies of the above mentioned cereal grains) of 98.6 and 99.3% were achieved when the morphology-texture model with the 15 most significant features was used to test on an independent data set (total number of kernels used was 10,500) and on the training data set (total number of kernels used was 31,500), respectively. The highest classification accuracies were achieved when the morphology-texture-color model was used.

Krasnogor and Smith (2005) studied several work done on memetic algorithms with the purpose of designing a syntactical model for MAs to enable a better understanding of the interplay between the different components of an MA. MAs are combination of evolutionary algorithm with local search. They constructed taxonomy of MAs to provide a conceptual framework which suggest areas of future research and facilitates algorithmic comparisons. This model suggested the existence of a novel class of metaheuristic in which several schedulers interact.

Kotsiantis, (2007) categorized various supervised machine learning classification techniques like logic-based learning techniques (decision trees, rule learners), perceptron-based techniques (neural network), statistical techniques (bayesian network, k-NN), and support vector machines (SVMs). He asserts

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

that while dealing with an specific application domain, the important issue is not whether a learning algorithm is superior to others, but under which conditions a particular method can significantly outperform others on a given application problem. A comparative performance of these learning techniques was presented and concluded that for multi-dimensional and continuous features, SVM and NN performs well while for discrete and categorical features logic-based systems performs better. Further, in SVM and NN a large sample size is significant for improving the classification accuracy whereas for a relatively small dataset Naïve Bayes may perform better. He also demonstrates how irrelevant features may result in the decay of the performance in case of k-NN and NN drastically. Eventually makes them inefficient. Finally he investigated the improvement in performance by creating ensemble of classifiers.

Lu and Weng, (2007) studied various image classification methods and techniques for improving the classification performance. They demonstrated some important issues like the availability of high-quality remotely sensed imagery and ancillary data, the design of a proper classification procedure, and the analyst's skills and experiences that affects the performance of the classifier and suggested that a suitable classification system and sufficient number of training samples are significant for improving classification accuracy.

Wang and Yong, (2008) proposed a texture analysis and classification approach with the linear regression model based on the wavelet transform. This method is motivated by the observation that there exists a distinctive correlation between the samples images, belonging to the same kind of texture, at different frequency regions obtained by 2-D wavelet packet transform.

Kurian and Karunakaran, (2012) performed another survey on image classification methods. They considered blurry and noisy images and compared the classification rate of various image clasiification methods like Bags of regions, Linear discriminant analysis (LDA), SVM, ANN, Bayesian classifier and Self organizing tree algorithm. They concluded that the best method among them for image classification was self organizing tree algorithm which is an unsupervised classification method.

FEATURE SUBSET SELECTION

Feature subset selection is an important issue in the research fields such as system modelling, data mining and pattern recognition. FSS evaluates a subset of features as a group for suitability prior to applying a learning algorithm. FSS algorithms can be broken into wrapper, filter and embedded. There are a number of domains where filter and wrapper approaches are applied for feature selection by various researchers. A chronological review of these methods is being presented.

Liu and Motoda, (1998) wrote their book on feature selection which offers an overview of the methods developed since the 1970s and provides a general framework in order to examine these methods and categorize them. This book demonstrated the importance of feature selection algorithms by employing various simple examples and compared those using different datasets. Also, there are demonstrations in this book for using different feature selection algorithms under various circumstances.

Filter Approach

Filter approach uses a search criterion that is independent of any learning algorithm to find the relevant feature subset. The goodness of a feature subset can be assessed using the intrinsic properties of the data.

Battiti, (1994) performed feature subset selection using mutual information criterion. Mutual information evaluates the "information content" of each individual feature with regard to the output class. The features are selected in a greedy manner, ranking them according to their MI with respect to the output class and with respect to the already-selected features into account. The approach was proved to be satisfactory in different classification areas like sonar target, iris, optical character recognition etc.

Lanzi, (1997) proposed a fast filter approach to feature subset selection using GA. This approach was independent of a specific learning algorithm and computationally efficient. It evaluated the fitness of individuals in the population using inconsistency rate. According to the experimental results this approach speeds up the feature selection process without any loss of predictive accuracy.

Hall, (1999) introduced a correlation based approach for feature selection. The central idea of the approach is highly correlated features with the class yet uncorrelaed with each other forms a good feature set. Correlation based feature selection (CFS) quickly removes irrelevant and redundant features and identifies relevant features. He further compared CFS with a wrapper approach and in many cases CFS gave comparable results to the wrapper approach.

Yu and Liu, (2003) introduced a filter approach based on correlation. It was a fast correlation based filter solution to identify relevant features as well as remove redundancy. They develop an algorithm FCBF (fast correlation based filter) to deal with high dimensional data. The experimental study suggested that FCBF not only enhanced the classification accuracy by selecting relevant features but also achieve high degree of dimensionality reduction.

Peng *et al.*, (2005) proposed a feature subset selection approach on mutual information criteria of max-dependency, max-relevance and min-redundancy. They derived an equivalent form to maximal dependency condition called mRMR (minimal-redundancy-maximal-relevance) criteria for first-order incremental feature selection. They also combined mRMR with other feature selectors like wrapper to select best features at very low cost. According to the experimental results, mRMR improves feature selection and classification accuracy on three different classifiers (Naïve Bayes, SVM, Linear Discriminant Analysis).

Sun *et al.*, (2005) presented a systematic Evolutionary Gabor Filter Optimization (EGFO) approach for on road vehicle detection that yields a more optimal problem-specific set of filters. EGFO approach unifies filter design with filter selection by integrating genetic algorithms (GAs) with an incremental clustering approach. The resulting filters were evaluated using an application oriented fitness criterion based on SVM.

Estévez et al., (2009) presented a filter method for feature selection based on mutual information, called normalized mutual information feature selection (NMIFS). NMIFS is an enhancement over Battiti's **(Battiti, 1994)** MIFS, MIFS-U, and mRMR methods. NMIFS outperformed MIFS, MIFS-U, and mRMR on several artificial and benchmark data sets without using any user-defined parameter. They introduced the normalized MI as a measure of redundancy, in order to reduce the bias of MI toward multivalued attributes and restrict its value to an interval. Further, NMIFS is combined with a genetic algorithm to form a hybrid filter/wrapper method called GAMIFS. GAMIFS overcomes the limitations of incremental search algorithms that are unable to find dependencies between groups of features.

Wrapper Approach

Wrapper approach uses the induction algorithm as a part of the evaluation function, the very algorithm that will be used to induce the final classification model.

Kohavi and John, (1997) is an influential paper which compares the wrapper approach to induction without feature subset selection and to Relief (a filter approach) to FSS. They present a number of disadvantages of the filter approach steering emphasis towards algorithms adopting the wrapper approach. Their approach search for an optimal feature subset tailored to a particular learning algorithm and a particular training set.

Dash and Liu, (1997) gave a survey of feature selection methods for classification. They categorized various feature selection methods on the basis of the generation procedure and evaluation function used. They described five types of evaluation functions and three types of generation procedures. Further they proposed a framework in which a total of 32 methods were categorized on the criteria of generation procedure and evaluation function. An entire set of 15 categories were presented for filter and wrapper based methods.

Yang and Honavar, (1998) used Genetic Algorithm for feature subset selection in automated design of pattern classifiers. They developed a simple, inter-pattern distance based provably convergent, polynomial time constructive neural network algorithm which compares favourably with computationally far more expensive algorithms in terms of generalization accuracy.

Inza *et al.*, (2000) proposed a new method for feature subset selection in machine learning called FSS-EBNA (FSS by estimation of Bayesian Network Algorithm). It utilizes the wrapper approach. It needed a large amount of CPU time with the ID3 learning algorithm but very less CPU time with Bayesian networks.

Inza *et al.*, (2002) published a paper proposing gene selection by sequential search wrapper approaches in microarray cancer class prediction. In this paper the gene selection process is performed by a sequential search engine, evaluating the goodness of each gene subset by a wrapper approach which uses leave one out process as the evaluation function.

Dy and Brodley, (2004) explored the problem of FSS, clustering and order identification for unsupervised learning within the wrapper framework. They addressed the issues in developing a feature subset

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

selection algorithm for unlabeled data. They used Expectation-Maximization clustering for Feature Subset Selection (FSSEM) and explored the issues associated with the development of algorithms under wrapper framework. The feature subset were evaluated on two different performance criteria namely scatter separability and maximum likelihood.

Hashemi, (2005) introduced a linear time wrapper to identify atypical points to further improve the performance of wrapper models. Atypical points are the data instances not useful for the classification task and often misclassified. The approach was shown to be 75 times more accurate than quadratic time wrapper. He tested the linear time wrapper over 7 classifiers and 20 data sets. He also proposed an algorithm Atypical Sequential Removing (ASR) which can eliminate atypical points without damaging the prediction accuracy in the data set.

Maldonado and Weber, (2009) proposed a wrapper method for feature selection using support vector machine. They introduced a novel wrapper algorithm for feature selection using SVM with kernel functions based on a sequential backward selection strategy. In sequential backward selection, features are removed one by one in subsequent iteration. They also compared their approach with a filter approach and a Recursive Feature Elimination SVM to demonstrate its effectiveness and efficiency.

Gutlein *et al.*, (2009) introduced a linear forward selection technique to reduce the number of attributes expansions in each step to overcome the problem of over fitting and high runtime that occurs mostly in wrapper approach. The technique showed good results as it is faster, finds even small subsets and accuracy is increased in comparison to standard forward selection. They also proposed a variant to this technique which eliminates the problem of overfitting by determining the subset size explicitly. The experimental result showed that the subset size is reduced significantly without deteriorating the accuracy.

Han *et al.*, (2011) demonstrated that wrapper type semi supervised feature selection method (FW-SemiFS) do not consider the confidence of predicted unlabeled data, rather evaluates the relevance of features according to their frequency. Therefore approach is computationally expensive. Hence, they proposed an ensemble-based semi-supervised feature selection method known as (EN-SemiFS) that considers the confidence of predictive unlabeled data. The experimental results confirmed that the proposed approach was faster and more accurate and able to select a more relevant feature subset using confident unlabelled data.

Shiue *et al.*, (2012) developed an ensemble of classifier based on GA wrapper feature selection approach for real time scheduling (RTS). It provides better system performance under all performance criterions. It enhanced the generalization ability of the classifier. They compared their approach with three classical machine learning-based classifiers, including the GA-based wrapper feature selection mechanism.

CONCLUSION

This paper gives a brief overview about Methods for 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 FEATURE SUBSET SELECTION.

REFERENCES

[1] Aeberhard, S.; Coomans, D. & De Vel, O. 1992. Comparison of classifiers in high dimensional settings. Dept. Math. Statist., James Cook Univ., North Queensland, Australia, Tech. Rep, (92-02).
[2] Bache, K. & Lichman, M. 2013. http://archive.ics.uci.edu/ml, UCI Machine Learning Repository, Irvine, CA: University of California, School of Information and Computer Science. Last visited 30th June.
[3] Barati, E.; Saraee, M.; Mohammadi, A.; Adibi, N. and Ahamadzadeh, M. R. 2011. A Survey on Utilization of Data Mining Approaches for Dermatological (Skin) Diseases Prediction in Cyber Journals: Multidisciplinary Journals in Science and Technology, JSHI, Volume 2, Issue (3), pp. 1-11.
[4] Battiti, R. 1994. Using mutual information for selecting features in supervised neural net learning. Neural Networks, IEEE Transactions on, Volume 5 Issue (4), pp.537-550.
[5] Braspenning, P. J.; Thuijsman, F. & Weijters, A. J. M. M. (Eds.). 1995. Artificial neural networks:an introduction to ANN theory and practice, Volume 931. Springer-Verlag.
[6] Cervantes, J.; Li, X. & Yu, W. 2008. Multi-Class SVM for Large Data Sets Considering Modelsof Classes Distribution. In

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

International Conference on DataMining(DMIN), pp. 30-35, CSREA Press, Las Vegas.

[7] Charytanowicz, M.; Niewczas, J.; Kowalski, P. A.; Kulczycki, P.; Łukasik, S. and Zak, S. 2010. Complete Gradient Clustering Algorithm for Features Analysis of X-Ray Images. Information Technologies in Biomedicine Advances in Intelligent and Soft Computing Volume 69, Issue (1), pp 15-24

[8] Chipperfield, A.; Fleming, P.; Pohlheim, H. & Fonseca, C. 1994. Genetic algorithm toolbox for use with MATLAB. University of Sheffield, Version, 1.

[9] Cortes, C. & Vapnik, V., 1995. Support-vector networks. Machine learning, Volume 20, Issue (3), pp. 273-297.

[10] Cover, T.; Thomas, J.; Proakis, J. G.; Salehi, M. & Morelos-Zaragoza, R. H. 1991.Elements of information theory. telecommunications. Wiley series.

[11] 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.